Statistical Inference for Autoregressive Models – with Random Coefficients and with Functional Realizations

Von der
Carl-Friedrich-Gauß-Fakultät
der
Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von
Thorsten Matthias Fink
geboren am 12.09.1982 in Radolfzell

Eingereicht am: 01.02.2013
Disputation am: 05.04.2013.

Referent: Prof. Dr. Jens-Peter Kreiß, Technische Universität Braunschweig
Koreferent: Prof. Dr. Efstathios Paparoditis, University of Cyprus

(2013)

Abstract

In the first part of this thesis (Chapters 2 through 5) we thoroughly examine autoregressive (AR) processes with random coefficients. We propose a least-squares estimator for the fourth order moments of both the noise sequences and state its consistency. The main theme is the development of various bootstrap procedures for the distribution of the autoregressive parameter and the distribution of the variances of both noise sequences. We show how to obtain approximative residuals for the process even though the standard method for autoregressive processes does not work in this context since one then would obtain convoluted residuals of both the noise squences. These ideas lead to a modification of the classical residual bootstrap for autoregressive processes.
First, the consistency of a bootstrap procedure for the autoregressive parameter, that is based on an intuitive least-squares estimator, is established. Further, the estimators proposed at the beginning are used to form two wild bootstrap modifications and the performances of the three bootstrap procedures are explored by a simulation study and compared to each other.
To overcome the drawback of very strong moment conditions that are imposed on the process to show consistency of these bootstrap procedures, the residual bootstrap is reconsidered using a quasi maximum likelihood estimator. Thereafter, as a variation of the residual bootstrap, we propose a wild bootstrap that uses estimated densities of the innovation and the disturbance noise to generate a bootstrap replicate of the process. Except some regularity conditions on the noise sequences, no moment assumptions on the process itself are needed to show consistency of both the bootstrap procedures for the distribution of the autoregressive parameter and the distribution of the variances of both the noise sequences. Their performance is illustrated by a simulation study.
Finally, we propose two basic estimators and an advanced estimator that is based on deconvolution techniques for the densities of the noise sequences. After some remarks concerning their practicability, we establish their consistency and evaluate their finite sample behavior by a simulation study.
In the second part of this thesis we consider functional time series that are assumed to follow an autoregressive scheme of unknown order and show how to estimate this order consistently. We precisely establish the connection between functional AR processes and multivariate AR processes and show how to obtain a multivariate process if we are given a functional AR process. The resulting process follows an autoregressive scheme, but is not a standard AR process anymore. The coefficient matrices are random and the residuals are dependent on the observations of the process. Following earlier contributions for AR processes, we introduce a general loss function and show that the estimated order obtained by a minimization of this function converges to the correct order of the multivariate non-standard AR process and therefore of the functional AR process in probability. We evaluate the finite sample size performance of this estimator by a simulation study and compare it with an existing method. Finally, we apply the method to real data sets.

Zusammenfassung

Im ersten Teil dieser Arbeit (Kapitel 2 bis 5) behandeln wir ausführlich autoregressive (AR) Prozesse mit zufälligen Koeffizienten. Wir schlagen einen kleinste-Quadrate-Schätzer für die vierten Momente beider Störgrößen vor und zeigen, dass er konsistent ist. Das Hauptziel ist die Entwicklung verschiedene Bootstrapideen, sowohl für die Verteilung des autoregressiven Parameters als auch der Varianzen der beiden Störgrößen. Wir zeigen, wie man approximative Residuen des Prozesses erhalten kann, obwohl die Standardmethode für AR Prozesse hier nicht funktioniert, da man dann nur Residuen erhalten würde, die aus der Summe beider Störgrößen bestehen. Diese Ideen führen zu einer Modifikation des klassischen Residuenbootstrap von AR Prozessen.
Zunächst zeigen wir die Konsistenz eines Bootstrapverfahrens für den AR Parameter, das auf einem kleinste-Quadrate-Schätzer basiert. Außerdem werden die am Anfang hergleiteten Schätzer verwendet, um zwei Wild-Bootstrapvarianten dieses Verfahrens herzuleiten. Schließlich wird das Verhalten der drei Bootstrapverfahren anhand einer Simulationsstudie untersucht und untereinander verglichen.
Um die starken Momentenannahmen zu umgehen, die nötig waren, um die Konsistenz dieser Verfahren zu zeigen, wird im folgenden Kapitel ein Quasi-Maximum-Likelihood-Schätzer für das Residuenbootstrap verwendet. Außerdem wird ein Wild-Bootstrapverfahren vorgeschlagen, das geschätzte Dichten der Störgrößen verwendet. Außer einigen Regularitätsannahmen an die Störgrößen werden nun keine Momentenannahmen an den Prozess selbst gestellt, um die Konsistenz beider Bootstrapverfahren für die Verteilung des AR Parameters und der Varianzen der Störgrößen herzuleiten. Das Verhalten der Verfahren wird anhand einer Simulationsstudie untersucht.
Schließlich entwickeln wir zwei einfache Schätzer für die Dichten der Störgrößen und einen, der auf Dekonvolutionstechniken beruht. Wir leiten ihre Konsistenz her und illustrieren ihr Verhalten mithilfe einer Simulationsstudie.
Im zweiten Teil dieser Arbeit betrachten wir funktionale Zeitreihen, die einem autoregressiven Schema unbekannter Ordnung entstammen und zeigen, wie diese Ordnung konsistent geschätzt werden kann. Dazu stellen wir präzise die Verbindung zwischen funktionalen AR Prozessen und multivariaten AR Prozessen her und zeigen, wie man aus einem funktionalen AR Prozess einen multivariaten Prozess erhalten kann. Dieser Prozess folgt ebenfalls einem autoregressiven Schema, ist allerdings kein Standard-AR-Prozess mehr. Die Koeffizientenmatrizen sind zufällig und die Residuen hängen von den Realisierungen des Prozesses ab. Früheren Ansätzen für AR Prozesse folgend, führen wir eine allgemeine Verlustfunktion ein und zeigen, dass die durch eine Minimierung dieser Funktion geschätzte Ordnung in Wahrscheinlichkeit gegen die richtige Ordnung des multivariaten AR Prozesses und damit des funktionalen AR Prozesses konvergiert. Wir untersuchen außerdem das Verhalten der Methode für endliche Stichprobenumfänge in einer Simulationsstudie und vergleichen sie mit einer existierenden Methode. Schließlich wenden wir die Methode auf reale Datensätze an.

Contents

List of Figures

List of Tables

1 Introduction

For many years, people are tracking measurements over time: river flow data, temperature, macroeconomic data, stock prices, to name just a few. The examples of these time series are numerous and the data basis and the measurement intervals vary. The hight of the nile river, for example, is documented starting in the year 622 already and measured annually. The unemployment rates are announced monthly and stock prices are computed nearly continuously. Nearly everybody comes into contact with time series in every day life. In general, a time series is just a sequence of some measurements that is collected in discrete time. In the classical sense, these measurements are real numbers or vectors of real numbers, but one can also think of a whole function that is obtained as one measurement. One could, for example, consider just the temperature at noon or one could consider the temperature curve of the whole day as one measurements. While the first one is often refered to as a time series the latter one is refered to as a functional time series.
The purpose of collecting data of a time series can be documentation, but usually people want to extract meaningful statistics and characteristics from the data and probably also make predictions and conclusions about the future development of the time series. Therefore, people are interested in the interdependency structure between the data collected at different points in time. The main purpose of time series analysis is to describe the data by models and so be able to determine some possible future events.
The basic underlying assumption is that one can measure the data at equidistant points in time such that one obtains realizations $X_1, \ldots, X_n$ of a random stochastic process $(X_t)_{t \in \mathbb{Z}}$ on some probability space $(\Omega, \mathcal{F}, P)$. This process, also referred to as time series, can be real valued, vector valued or also function valued. The goal is to gather as much information as possible about the unknown process $(X_t)_{t \in \mathbb{Z}}$ from the observations $X_1, \ldots, X_n$.
An extensive overview about time series analysis can be found in Kreiss & Neuhaus (2006) and Brockwell & Davis (1991). In the following, we will shortly review the most important facts, present the proceeding in analyzing a given data set, and especially elaborate the models and concepts that we will consider in more detail in Chapters 2 through 6 of this thesis.

1.1 Overview of time series analysis

Basic concepts

A widly used assumptions, that we will also ask for in the following, is the presence of stationarity in the data and thus in the process $(X_t)_{t \in \mathbb{Z}}$:

Definition 1.1.1. *A process $(X_t)_{t \in \mathbb{Z}}$ is said to be strictly stationary if for all $n \in \mathbb{N}$ and all $t_1, \ldots, t_n$, $h \in \mathbb{Z}$ holds that $\mathcal{L}(X_{t_1}, \ldots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \ldots, X_{t_n+h})$.*

If $\boldsymbol{E}[X_t^2] < \infty \; \forall t$ *it is said to be weakly stationary, covariance stationary, or simply stationary if* $\boldsymbol{E}[X_t] = \mu$ *and* $\boldsymbol{Cov}[X_{t+h}, X_t] = \gamma(h)$ *are both independent from* t *for all* $t, h \in \mathbb{Z}$.

This means, that for a stationary process certain distributional properties do not change over time, it exhibits basically the same characteristics over time. In reality, data often cannot be considered as stationary, because it exhibits trends and seasonal behavior. Thus, one has to account for these effects first and has to eliminate these ratios to obtain a stationary time series before further analysis can be performed. One famous example are the unemployment rates that are cited in the news as for seasonal influence adjusted numbers. While the assumption of weak stationarity is very common and much easier to check than the presence of strict stationarity, there are a lot of examples of data, such as financial data, and models, some of which we will consider further later, that are heavy-tailed and so do not have finte second moments. In the discussion now following we will restrict ourselves for notational convenience to real valued time series, all results can be generalized to vector valued time series easily. A basic example of a stationary time series is a white noise:

Definition 1.1.2. *A real or complex valued sequence* $(e_t)_{t\in\mathbb{Z}}$ *is called white noise if* $\boldsymbol{E}[e_t] = 0$ *and* $\boldsymbol{E}[|e_t|^2] = \sigma^2 \in (0, \infty)$ *for all* $t \in \mathbb{Z}$ *and* $\boldsymbol{Cov}[e_{t_1}, e_{t_2}] = 0$ *for all* $t_1 \neq t_2 \in \mathbb{Z}$.

Remark 1.1.3. *In its general definition, a white noise is a weakly stationary time series. But frequently the assumption that the random variables* $(e_t)_{t\in\mathbb{Z}}$ *are independent and identically distributed (i.i.d.) is added so that it becomes a strictly stationary time series as well. Later on, we will also consider i.i.d. white noises.*

Linear Processes

However, the assumption of stationarity itself is important but usually not enough to state important results such as central limit theorems. Therefore, one has to ask for additional assumptions on the time series, such as linearity or dependence conditions like mixing or weak dependence. Both are concepts to model the dependence between two different points in time of a time series that in the context of mixing or weak dependence decreases with the time gap between these two processes increasing. Depending on the rates at which this dependence decreases the processes can be classified into different classes of mixing processes, see, for example, Dedecker et al. (2007) for an overview. From the white noise the famous linear processes that include autoregressive processes and moving average processes as special cases can easily be deduced.

Definition 1.1.4. *A stochastic process* $(X_t)_{t\in\mathbb{Z}}$ *is called linear process if it exhibits a representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \qquad t \in \mathbb{Z},$$

where $(b_j)_{j\in\mathbb{Z}}$ *with* $b_0 = 1$ *is an absolutely summable real valued sequence and* $(e_t)_{t\in\mathbb{Z}}$ *is an i.i.d. white noise. If* $b_j = 0$ *for all* $j < 0$, $(X_t)_{t\in\mathbb{Z}}$ *is called causal linear time series.*

Deduced from these processes can with a white noise $(e_t)_{t\in\mathbb{Z}}$ the causal moving average processes of order q (short MA(q) processes)

$$X_t = \sum_{j=0}^{q} b_j e_{t-j}, \qquad t \in \mathbb{Z},$$

and the autoregressive processes of order p (short AR(p) processes) that are solutions of the equation

$$X_t - \sum_{i=0}^{p} a_i X_{t-i} = e_t, \qquad t \in \mathbb{Z},$$

as well as the combination of both, the autoregressive moving average processes of orders p and q (ARMA(p,q) processes), that are solutions of the equation

$$X_t - \sum_{i=0}^{p} a_i X_{t-i} = \sum_{j=0}^{q} b_j e_{t-j}, \qquad t \in \mathbb{Z}.$$

Remark 1.1.5. *The orders $p, q = \infty$ are allowed as well. Under additional assumptions on the roots of the corresponding polynomial $A(z) = \sum_{i=0}^{p} a_j z^j$ the AR(p) and the ARMA(p, q) process is stationary. Under suitable conditions, an AR(p) process can be written as an MA(∞) process. This representation is very useful and widely used. The process is called causal if this representation is of the form*

$$X_t = \sum_{j=0}^{\infty} a_j e_{t-j},$$

that means X_t is only dependent on e_s with $s \leqslant t$. Stationary and causal AR(p) processes are mixing and weakly dependent.

Non-linear processes

Linear processes exhibit some desirable features, allow for a large variety of processes that can be modelled, and are statistically (relatively) easily tractable. However, the assumption that X_t depends linearly on $(e_t)_{t\in\mathbb{Z}}$ limits the possible characteristics at some extent. Financial data, for example, exhibits heteroscedasticity and biological data suffers from random perturbations. A generalization are non-linear time series. They can represent more properties, but are statistically harder to capture. Two famous models are the ARCH and GARCH model that attracted a lot of attention since Robert Engle received the Nobel Prize in Economic Science in 2003 for introducing these models. They do not only model the process itself, but also model the volatility depending on the past values of the volatility and possibly the past values of the process. They are especially famous for financial time series because they allow for heteroscedascity and volatility clustering, two frequently observed characteristics that cannot be modelled with linear time series. Another obvious generalization of a linear process or an autoregressive process is to model the coefficients stochastically, that means adding a (typically relatively small) noise term to the coefficients, such that one arrives at the so-called random coefficient autoregressive model:

Definition 1.1.6. *A Random Coefficient Autoregressive Process of order* p *(RCA Process) is given as the solution of the equation*

$$X_t = \sum_{i=1}^{p} (\varphi_i + b_{t,i})\, X_{t-i} + e_t, \qquad t \in \mathbb{Z}$$

with two white noises $(b_t)_{t\in\mathbb{Z}}$ *and* $(e_t)_{t\in\mathbb{Z}}$ *.*

We will restate this Definition in a slightly restricted form for order $p = 1$ in Chapter 2. In contrast to the GARCH model where the volatility is modelled as a non-linear process the RCA process varies the process directly. It also allows for behavior that looks at first sight as unstationarity, such as periods of high and low volatility or large outliers, what is freqently observed in time series. Originally, these processes were introduced to investigate random perturbations of dynamical systems but by now a variety of applications, for example, in finance and biology can be found (see Tong (1990) for an overview). These models allow to some extent for the same statistical methods as the standard autoregressive model, but also demand for advanced methods. The models were first introduced by Andel (1976) and comprehensively studied by Nicholls & Quinn (1980) and Nicholls & Quinn (1982). Other authors include Schick (1996), Koul & Schick (1996), and Berkes et al. (2009). Recent studies were conducted by Aue et al. (2006) who also considered several other cases, like, for example, the non-stationary case or the parameter stability (Aue (2006), Aue & Horvath (2011)).

Order selection of AR processes

Having decided a model of which class should be fitted to a data set, the next question is, how complex the concrete model should be, meaning how many parameters should be included in the model, for example how large should be the order p of an AR model and thus how many parameters $a_1, \ldots, a_p$ should be included in the model.
There is a vast literature on univariate and multivariate autoregressive processes, including order selection, see Lütkepohl (2005) for an extensive overview on VAR-processes, including statistical tools for their examination and order selection. Earlier contributions on order selection for AR-processes include the criteria AIC (Akaike (1974)), AICC (Hurvich & Tsai (1989)), BIC (Akaike (1978)), FPE (Akaike (1969)), and HQ (Hannan & Quinn (1979)), multivariate extensions of these methods can be found in Lütkepohl (2005), Reinsel (2003), Tiao & Tsay (1989) and Quinn (1980). Another famous criterion for order selection can be deducted from the MDL criterion as introduced by Rissanen (1989) and Rissanen (2007), see also Lee (2001) for a tutorial introduction to this method for selecting a best fitting model from a class of candidate models. They are all based on the same principle, namely that for various possible choices of the order the values of a loss function taking into account the accuracy and the complexity of the model are compared.
Paulsen (1984) and some former work of the author considers the order selection for multivariate AR-processes in great detail and also includes the case that the time series has unit roots, while the loss functions are allowed to follow a general function and thus several of the criterions mentioned above are included, for example the MDL, but not the famous AIC.

Functional time series

When hearing of time series (analysis) one probably first thinks of real valued or vector valued time series as described before. But, as already indicated at the beginning, there is another huge class of processes: the functional time series. Although being a rich class of models with a broad range of applications it did not receive the same attention as the classical time series analysis yet. However, functional data analysis has established itself as an important and dynamic area of statistics in recent years. Functional data often arises from measurements on fine time grids. It is collected in sequential form and several curves are obtained by separating an almost continuous time record into consecutive intervals, for example days, weeks, or years, for that similar behavior is expected. This could include daily price and return curves of financial assets or the daily pattern of geophysical, meteorological and environmental data, but also annual temperature profile curves, see, for example, Berkes et al. (2009).
The resulting functions may be described by a functional time series $Y_t,\ t \in \mathbb{Z}$,

$$Y_t(\tau), \quad t \in \mathbb{Z},\ \tau \in [0, \mathcal{T}] \tag{1.1}$$

each term in the sequence being a random function $Y_t(\tau)$ defined for τ taking values in some interval $[0, \mathcal{T}]$. For this time series the assumption of independence is often too strong to be realistic in many applications. Especially if the data is collected sequentially over time, it is natural to expect that the current observation depends to some degree on the previous observations. Due to this, in analogue to standard time series, also for functional data appropriate functional time series models emerged. We refer to Ferraty & Vieu (2006), Ferraty & Romain (2011), and Ramsey & Silverman (2005) for a general introdruction into functional data analysis. Bosq (2000) covers the mathematical foundations and also introduces the functional AR process and thoroughly examinates the theoretical properties and characteristics of this process. This model became probably the most popular functional time series model so far.
A recent survey on time series aspects is given by Hörmann & Kokoszka (2012). They extensively study functional time series models and develop a theory to examine functional data, especially the concept of m-approxibility is crucial for this theory. They also consider functional principal component analysis for time series and the convergence of the estimators for the eigenfunctions and eigenvalues thereby extending the results for the independent case by Dauxois et al. (1982), Bosq (2000), Bosq & Blanke (2007), and Ramsey & Silverman (2005). We will use these results to transform a given functional time series into a multivariate standard time series thereby precisely establishing the connection between functional autoregressive processes of order p (FAR(p) processes) and vector autoregressive processes of order p (VAR(p) processes).
Concerning the functional autoregressive process, most theory is developed for FAR(1) processes, only one paper covering order selection is known to the author. Kokoszka & Reimherr (2012) propose a multistage testing procedure for different orders by representing the FAR(p)-process as a fully functional linear model with dependent regressors. However, there is some literature available on prediction of functional time series, see Aue et al. (2012) and the references therein. In contrast, as already mentioned, there is a broad range of literature on (multivariate) AR processes available and also order selection is well developed for these processes. We will generalize these results to FAR-processes.

1.2 Overview of bootstrap

Analyzing data and fitting a model to data always means dealing with uncertainty and randomness, meaning that one always has to estimate quantities from the data after having decided which model class and which concrete model should be fitted to the data. One very important question that comes to ones mind after having estimated some quantities is how reliable the estimates are. This means, how close are the estimators to the true (unknown) values and how do they change with a change in the data. To answer these questions, some asymptotic theory to construct for example confidence intervals for the true parameters has to be carried out. Under suitable assumptions one can, for example, state a central limit theorem and find a limiting distribution of the estimator, typically a normal distribution. Hence, one can approximate the unknown distribution of the statistic by a normal distribution. While this approach is relatively easy, it has some substantial drawbacks: The results are only of asymptotic nature and it is not clear how well their quality is in finite sample sizes. In addition, the normal distribution forces the asymptotic distribution to be symmetric even if it would be skewed in reality. Both of these drawbacks can the confidence intervals lead to be of poor quality and not realiable. Furthermore, often it is also not even possible to derive the asymptotic distribution or to determine the asymptotic variance out of the data because it might be necessary to estimate additional statistics what is not always possible or adds considerably more uncertainty again.
Over the last decades, resampling procedures emerged. One technique that is famous along statisticians and also practitioners is the bootstrap, that was introduced for i.i.d. random variables by Efron (1979). It has been acknowledged as a powerful tool for approximating certain distributional characteristics of statistics which are sometimes difficult to compute or even not possible to derive analytically.

Bootstrap Procedure 1.2.1. *Basically, the bootstrap is as follows:*

- *We are given a sample $X_1, \ldots, X_n$ of i.i.d. random variables and want to approximate a statistic $T_n = T_n(X_1, \ldots, X_n)$.*
- *We sample with replacement from the original sample $\{X_1, \ldots, X_n\}$ and obtain the so-called bootstrap sample $X_1^*, \ldots, X_n^*$.*
- *We compute the bootstrap statistics $T_n^* = T_n^*(X_1^*, \ldots, X_n^*)$.*
- *We repeat the last two steps frequently to determine the empirical distribution of T_n^* that approximates the distribution of T_n^*.*
- *If we can show that the distributions of T_n^* and of T_n are close together, we can see the one of T_n^* as an approximation of the distribution of T_n.*

In practice, one typlically shows that T_n and T_n^* have the same limiting distribution. While this way of proving consistency of the bootstrap does not give us any certainty that for finite samples the distributions of T_n and of T_n^* are closer together than the distribution of T_n and the limiting (normal) distribution, though in practice it can usually be seen that this is the case and that the bootstrap considerably outperforms the approximation via the normal distribution. The big advantage is that the bootstrap can account for possible

skewness in the distribution very well and can give results also if no limiting distribution of the statistic like a normal distribution exists or if it cannot be determined easily from the data.
However, the procedure just described is valid for i.i.d. random variabels only and does not generalize to dependent data as one typically obtains from time series in a straightforward way. Especially for time series, a variety of methods emerged, each of which being tailored for a specific situation and having drawbacks and advantages, see Härdle et al. (2003) or Lahiri (2003) and also Kreiss & Paparoditis (2011) or Kreiss & Lahiri (2012) for an overview. In general, four types can be classified:

- **Block bootstrap** (Carlstein (1986), Künsch (1989), Politis & Romano (1994), etc.): The basic idea is closely related to the i.i.d. boostrap. The sample of size n is divided in m blocks of length l and it is drawn with replacement from these blocks. Due to the strict stationarity of the time series the blocks are i.i.d. and deliver the i.i.d. requirement for the bootstrap. In the blocks the data is dependent and delivers the structure necessary to capture all dependencies. To show consistency, it is necessary, that the number of observations n, the number of blocks m and the length of the blocks l grow to infinity. This method is non-parametric and can therefore be used very generally but it often performs considerably less accurate than parametric approaches if these are also possible.

- **Residual bootstrap** (Freedman (1984), Efron & Tibshirani (1986), Kreiss (1988), Kreiss (1997), Bühlmann (1997), Kreiss et al. (2011), etc.): The idea is to fit a parametric model to the data and to perform the classical i.i.d. bootstrap on the estimated residuals that can be assumed to be approximately i.i.d.. These methods are tailored for a specific situation and usually show very good behavior in simulations.

- **Frequency bootstrap** (Franke & Härdle (1992), Dahlhaus & Janas (1996), Kreiss & Paparoditis (2003), Shao & Wu (2007), etc.): These bootstrap approaches rely on the asymptotic features of the periodogram. The periodogram evaluated at different frequencies is asymptotically independent and thus no parametric assumption is needed. Mostly, these methods show reasonable behavior in simulations but because they are set up in the frequency domain and not in the time domain, they are limited to statistics of the periodogram.

- **Hybrid boostrap** (Jentsch & Kreiss (2010), Kirch & Politis (2011), Kreiss & Paparoditis (2012), etc.): This is an advanced technique that wants to overcome the main drawbacks of the frequency and the residual bootstrap: the limitation to the frequency domain on the one hand and the parametric assumption on the other hand. The idea is to fit a parametric model to the data, for example an AR model, to capture the essential features of the data and thereafter to perform a non-parametric correction in the frequency domain to catch the features not represented by the parametric fit. These methods also show good results in simulations.

In the following, we want to introduce reasonable bootstrap approaches for the situation of a random coefficient autoregressive model and therefore follow the residual bootstrap further.

1.3 Main results of this thesis

In this section, we want to give an overview about the following chapters and summerize the main results. All chapters except the last one are concerned with RCA processes, the last chapter will consider FAR processes. Chapter 2 is concerned with parameter estimation of RCA processes. We thoroughly introduce RCA processes and examine their main characteristics. We also present existing results concerning estimation of the parameter φ and the variances of both the white noises. We then generalize these results to the estimation of fourth moments of both the white noises and conclude this chapter with a simulation study.

In Chapter 3 we present a bootstrap approach for RCA processes as the generalization of the classical residual bootstrap for AR processes. This approach works for the distribution of the parameter φ. The crucial step is to obtain estimated residuals since the classical idea of constructing estimated residuals $\hat{u}_t = X_t - \hat{\varphi} X_{t-1}$ would lead to convoluted residuals of both the white noises and the sum cannot be splitted up. Hence, we consider just "small" observations $|X_{t-1}| < \varepsilon = \varepsilon(n) \stackrel{n\to\infty}{\longrightarrow} 0$ to construct estimated residuals $\hat{e}_t$ and just "large" observations $|X_{t-1}| > M = M(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ to construct estimated residuals $\hat{b}_t$. After that, we establish the consistency of this bootstrap procedure and present two variations of it before we perform a simulation study to evaluate the finite sample behavior of the bootstrap approaches. Like in all following chapters, the proofs that need some more space can be found at the end of the chapter.

The bootstrap approach presented in Chapter 3 is manifest, but strong moment assumptions on the process are needed to show consistency. This means that the parameter space and the possible characteristics that can be modelled with these processes are very limited. Furthermore, the former approach does only work for the distribution of the parameter φ, to show consistency for the distribution of the variance of both the white noises even stronger moment assumptions on the process itself would be required. Therfore, we present another, more elaborated, approach in Chapter 4. For this approach, we only need some basic regularity conditions on the noise sequences, but no moment conditions on the process itself. In particular, the process does not have to have finite second moments and thus does not have to be weakly stationary. In addition, this approach can approximate the joint distribution of the parameter φ and the variances of both the white noises simultaneously. We present a residual boostrap and a wild bootstrap that uses estimated densities of the noise sequences. In showing the consistency of the bootstrap approaches the crucial step is to show that the bootstrap can mimic terms of the form

$$\mathbf{E}\left[\frac{X_{i-1}^{\kappa}}{\left(\omega^2 X_{i-1}^2 + \sigma^2\right)^{\gamma}}\right], \quad \gamma = 0, 1, 2, \quad \kappa = 0, \ldots, 2\gamma,$$

that determine the asymptotic variance of the estimator, correctly. We inspect the performance of the bootstrap approaches based on the QML estimator by a simulation study and like for the bootstrap approaches based on the LS estimator we find out that they perform considerably better than a normal approximation.

In Chapter 5 we precisely derive some estimators for the densities of both the white noises that are needed for the aforementioned bootstrap approach. The first attempt is based on the same idea as the residual based bootstrap in just considering the "small" or the "large" observations of the process to obtain estimated residuals for the two noise sequences and then to use a standard kernel density estimator to obtain density estimates. Because for this method it is not clear which observations are "large" and if there are enough observations available to obtain reliable results, another approach based on deconvolution methods is presented to obtain an estimate of the density of the sequence $(b_t)_{t\in\mathbb{Z}}$ and some variations for finite sample sizes are suggested. All of the estimators are carefully elaborated and their asymptotic behavior in form of central limit theorems is evaluated. In addition, an extensive simulation study is performed that elaborates the finite sample behavior, which variation outperforms and how the parameters like the bandwidths should be chosen.

Finally, in Chapter 6 we leave the RCA processes and the bootstrap context and turn to FAR processes and consider the problem of order determination for FAR(p) processes. We first give an overview about FAR processes and since they are statistically hard to deal with, we show how they can be transferred into multivariate processes with non-standard, that means dependent and not identically distributed, noise structure. We introduce a general loss function and show how a consistent estimator for the order p of the non-standard VAR and thus of the FAR process can be obtained by using this loss function. This is a generalization of existing methods for multivariate AR processes. The crucial step is the generalization to a noise sequence that exhibits interdependencies and various error terms that in part vanish with the number of observations growing to infinity and in part do not. The question further is, how the dimension d of the multivariate process should be chosen. We note that we do not need the dimension to grow to infinity for our method to be consistent. We conclude our considerations again by an extensive simulation study. Therefore, we specify the general loss function to the MDL criterion and also use the AIC criterion, that does not fall in the class of the general loss function introduced before but is famous with practitioners. We also introduce a method to choose the dimension d in finite sample sizes in a lucrative way. For various scenarios we elaborate the finite sample behavior of our procedure and compare it to the existing sequential testing method suggested by Kokoszka & Reimherr (2012). We find out that the rule to choose p and especially the combination of choosing p and d simultaneuosly performs very well. Finally, we analyze two real data sets, australian mortality data and austrian temperature data, suggest an order for them and check by an out-of-sample prediction the quality of the results that seem to be satisfactory.

Summing up, the main results of this thesis are the presentation of a bootstrap approach for RCA processes in Chapter 3, the generalization of this approach by weakening the assumptions in Chapter 4, giving some results concerning the density estimation for RCA processes in Chapter 5 and finally the suggestion of a new approach to determine the order of a FAR process in Chapter 6.

Parts of Chapter 2 and Chapter 3 are based on a paper that has been submitted to Journal of Time Series Analysis as *Fink, T. and Kreiß, J.-P. (2012): Bootstrap for Random Coefficient Autoregressive Models.*

Chapter 4 is based on a paper that is submitted to Journal of the Korean Statistical Society as *Fink, T. and Kreiß, J.-P. (2013): Simultaneous Bootstrap for all three Parameters in Random Coefficient Autoregressive Models.*

Chapter 6 is based on a paper that is planned to be submitted to Journal of the American Statistical Association as *Aue, A., Fink, T., and Lee, T. (2013): Order Determination of Functional Autoregressive Processes.*

2 Parameter estimation of random coefficient autoregressive processes

2.1 Characteristics of RCA processes

Like many other authors, we also just consider the random coefficient autoregressive model of first order for ease of calculations. In principle, our methods can be extended to the general case. Throughout this thesis we would like to use the following Definition of a Random Coefficient Autoregressive Model.

Definition 2.1.1. *A Random Coefficient Autoregressive Process of order 1 (RCA Process) is given as the solution of the equation*

$$X_t = (\varphi + b_t)\, X_{t-1} + e_t, \qquad t \in \mathbb{Z} \tag{2.1}$$

where $\varphi \in \mathbb{R}$ and $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t\in\mathbb{Z}}$ both are i.i.d. centered random variables with variance ω^2 or σ^2, fourth moment α^4 or β^4, and density $h(\cdot)$ or $k(\cdot)$, respectively. Further, we assume $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t=1\in\mathbb{Z}}$ to be mutually independent and the density $h(\cdot)$ to have a connected support.

Throughout Chapters 2 to 5 we assume X_t to be a RCA process and b_t and e_t as given above. We will refer to the white noise $(e_t)_{t\in\mathbb{Z}}$ as the ***innovation noise***, to the white noise $(b_t)_{t\in\mathbb{Z}}$ as the ***disturbance noise***, and to φ as the ***autoregressive parameter***. In some chapters we will also put the assumption on the process that the odd moments of both up to order seven are zero and the eight moments of both are finite. We also assume that we are given a stationary process, which is characterized as follows.
The conditions under which the RCA Equation (2.1) has (stationary) solutions were discussed frequently in the literature. First work was done by Nicholls & Quinn (1980) and Nicholls & Quinn (1982) for the univariate and the multivariate case. Recently, Aue et al. (2006) extended this work. If

$$\mathbf{E}\left[\max(\ln|e_0|, 0)\right] < \infty, \quad \mathbf{E}\left[\max(\ln|\varphi + b_0|, 0)\right] < \infty, \quad -\infty \leqslant \mathbf{E}\left[\ln|\varphi + b_0|\right] < 0 \tag{2.2}$$

the equation (2.1) has a strictly stationary, absolutely convergent, causal solution that is given by

$$X_t = \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i-1} (\varphi + b_{t-j}) \right) e_{t-i} \quad \text{a.s.} \tag{2.3}$$

For the moments of an RCA process we have the following result.

Theorem 2.1.2. *Under the assumptions of Definition 2.1.1 with the addition that the odd moments of both of the white noises up to order three for the fourth moment and up to order seven for the eighth moment are zero it holds true that*

$$\begin{aligned} \boldsymbol{E}\left[X_t^2\right] < \infty &\iff \varphi^2 + \omega^2 < 1, \\ \boldsymbol{E}\left[X_t^4\right] < \infty &\iff \varphi^4 + 6\varphi^2\omega^2 + \alpha^4 < 1, \\ \boldsymbol{E}\left[X_t^8\right] < \infty &\iff \varphi^8 + 28\varphi^6\omega^2 + 70\varphi^4\alpha^4 + 28\varphi^2\boldsymbol{E}\left[b_t^6\right] + \boldsymbol{E}\left[b_t^8\right] < 1. \end{aligned} \tag{2.4}$$

If the moments are finite, it holds

$$\boldsymbol{E}\left[X_t^2\right] = \frac{\sigma^2}{1-\varphi^2-\omega^2} \qquad \text{and} \qquad \boldsymbol{E}\left[X_t^4\right] = \frac{\beta^4 + 6\sigma^4\frac{\varphi^2+\omega^2}{1-\varphi^2-\omega^2}}{1-\varphi^4-6\varphi^2\omega^2-\alpha^4} \tag{2.5}$$

The autocovariance is for $h \geqslant 0$ given by

$$\begin{aligned} \boldsymbol{Cov}\left[X_t, X_{t+h}\right] &= \varphi^h \boldsymbol{E}[X_t^2] \\ \boldsymbol{Cov}\left[X_t^2, X_{t+h}^2\right] &= (\varphi^2+\omega^2)^h \left(\boldsymbol{E}[X_t^4] - \boldsymbol{E}[X_t^2]^2\right) \end{aligned}$$

Proof. The first assertion in (2.4) is proved in Nicholls & Quinn (1982), Corollary 2.3.2, the other assertions in (2.4) follow similarly. The assertions about the second moments in (2.5) are stated in Aue (2003), Lemma 3.1.1. For the fourth moment we note that we assumed the odd moments of the white noises up to order seven to be zero and consider the term

$$\begin{aligned} \mathbf{E}\left[X_t^4\right] &= \mathbf{E}\left[\left(\left(\varphi + b_t\right) X_{t-1} + e_t\right)^4\right] \\ &= \mathbf{E}\left[(\varphi + b_t)^4\right]\mathbf{E}\left[X_{t-1}^4\right] + 6\mathbf{E}\left[(\varphi+b_t)^2\right]\mathbf{E}\left[X_{t-1}^2\right]\mathbf{E}\left[e_t^2\right] + \mathbf{E}\left[e_t^4\right] \\ &= \left(\varphi^4 + 6\varphi^2\omega^2 + \alpha^4\right)\mathbf{E}\left[X_{t-1}^4\right] + 6\left(\varphi^2+\omega^2\right)\mathbf{E}\left[X_{t-1}^2\right]\sigma^2 + \beta^4. \end{aligned}$$

Since the process is assumed to be stationary and since its first and third moment have to be zero (c.f. calculations in Aue (2003), Lemma 3.1.1) this equation transforms into

$$\mathbf{E}\left[X_t^4\right]\left(1 - \left(\varphi^4 + 6\varphi^2\omega^2 + \alpha^4\right)\right) = 6\left(\varphi^2+\omega^2\right)\mathbf{E}\left[X_t^2\right]\sigma^2 + \beta^4$$

from which the result can be obtained by plugging in the expression for the second moment from (2.5). The result about the autocovariance of the process is also given in Aue (2003), Lemma 3.1.1, and the remaining result about the autocovariance of the squared process follows by similar computations. □

Remark 2.1.3. *For normally distributed b_t the last term of (2.4) transforms into*

$$\varphi^8 + 28\varphi^6\omega^2 + 210\varphi^4\left(\omega^2\right)^2 + 420\varphi^2\left(\omega^2\right)^3 + 105\left(\omega^2\right)^4 < 1.$$

Remark 2.1.4. *Definition 2.1.1 together with Equation (2.3) immediately implies that in the context considered here a stationary RCA process X_t has a density as well. We denote this density by $f(\cdot)$ and the corresponding cumulative distribution function by F.* □

We want to illustrate under which conditions a strictly stationary solution to Equation (2.1) exists and when the process has finite moments of at least order two. Figure 2.2 (left) shows how the parameters φ and ω^2 can be chosen for the process to have finite moments of second, fourth, or eighth order, respectively. It can be seen that the parameter space is very limited for the finite second moment already and is extremly limited for the finite eighth moment. However, as the following discussion will show, the parameters can be chosen much more flexible, if one abstains from any moment conditions and just asks for a strictly stationary solution of Equation (2.1) as given by the condition stated in Equation (2.2):

$$1 > \prod_{j=0}^{i-1}(\varphi + b_{t-j}) \Leftrightarrow \quad 0 > \frac{1}{i}\ln\left|\prod_{j=0}^{i-1}(\varphi + b_{t-j})\right| = \frac{1}{i}\sum_{j=0}^{i-1}\ln|\varphi + b_{t-j}| \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}\left[\ln|\varphi + b_1|\right].$$

It further holds that

$$\varphi^2 + \omega^2 < 1 \;\Rightarrow\; (\mathbf{E}\left[\varphi + b_1\right])^2 < 1 \;\Leftrightarrow\; (\mathbf{E}\left[|\varphi + b_1|\right])^2 < 1 \;\Rightarrow\; \mathbf{E}\left[\ln|\varphi + b_1|\right] < 0$$

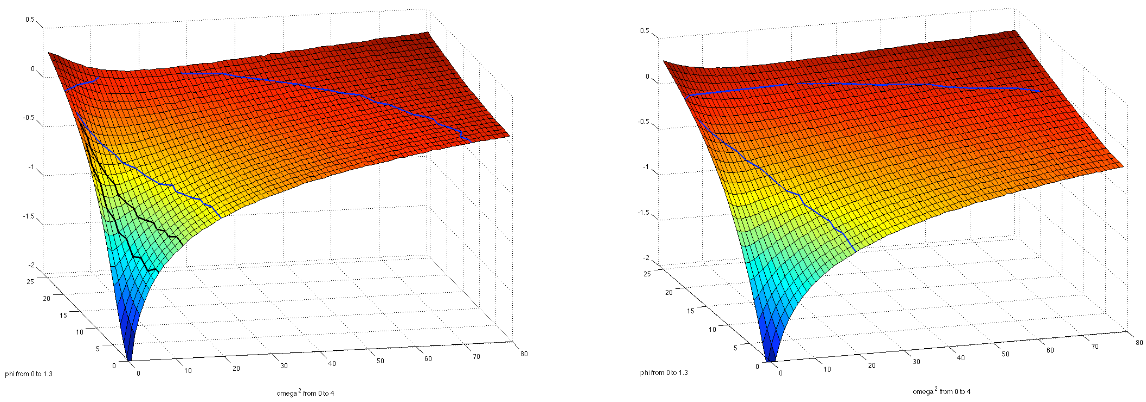

Figure 2.1: Simulation of $\mathbf{E}\left[\ln|\varphi + b_1|\right]$ with b_1 normally (left) and double-exponentially distributed (right)

This shows that stationary solutions X_t of Equation (2.1) without finite second moments can exist, that means that these processes are strictly stationary but not weakly stationary. How sharp are the bounds? To illustrate this, we perform some simulations of the expression $\mathbf{E}\left[\ln|\varphi + b_1|\right]$ with different parameters φ and ω^2 and with b_1 following a normal and a double-exponential distribution. The results are displayed in Figure 2.1. To the right ω^2 is increasing from 0 to 4, to the back φ is increasing from 0 to 1.3. The blue line in the front displays the values of φ and ω^2 with $\varphi^2 + \omega^2 = 1$. The blue line in the back marks the end of the zone with the empirical mean of $\ln|\varphi + b_1|$ being smaller than one. The black lines display those values of φ and ω^2 that divide the parameters into the parameters leading to processes that do have finite fourth moments and those that do not.
The two blue lines differ substantially from each other. For large φ a medium variance ω^2 is allowed while a variance ω^2 that is too small or too large does not lead to stationary

solutions. If φ is small, the variance ω^2 can grow enormously. The double-exponential distribution allows for a larger variance ω^2 than the normal distribution for small and median values of φ but does not allow for such large values for φ than the normal distribution.
Interestingly, φ can be greater than one for a small range of ω^2 indicating that here is no form of the well-known unit root problem that exists for standard autoregressive processes. This means, that already adding a noise with very small variance to the autoregressive paramter $\varphi \approx 1$ can circumvent the problems that are encountered with the unit root problem and can lead to a strictly stationary process, though probably without finite second moments.
However, a great variety of processes can be modelled with RCA processes. For small variance ω^2 of the disturbance noise, the processes look very much like standard autoregressive processes, wheras for large variance ω^2 very large peaks in the paths of the process can be produced with a stationary process. This admits for a very broad range of applications. Some sample paths of these different processes are shown in Figures 2.2 (right) and 2.3. Figure 2.2 (right) shows a sample path of a strictly stationary RCA process without finite moments and parameters $(\varphi, \omega^2, \sigma^2) = (0.95, 0.65, 0.8)$. The paths exhibit very large peaks but return to zero as well. Figure 2.3 shows a sample path of a stationary RCA process with finite second moments and parameters $(\varphi, \omega^2, \sigma^2) = (0.55, 0.6, 0.8)$ (left) and of a process with finite eighth moments and parameters $(\varphi, \omega^2, \sigma^2) = (0.55, 0.145, 0.8)$ (right). Having finite second moments, large peaks can appear, but the process mostly fluctuates around zero, having finite eighth moments, the process looks very much like a standard autoregressive process. We will consider these two stationary processes from time to time again during the following chapters. We will also refer to $(\varphi, \omega^2, \sigma^2) = (0.55, 0.145, 0.8)$ as *parameter set I* and to $(\varphi, \omega^2, \sigma^2) = (0.55, 0.6, 0.8)$ as *parameter set II*. For all of the processes, we choose the innovation noise to be double-exponentially distributed and the disturbance noise to be normally distributed. The reason for choosing these two distributions will become obvious in Chapter 5 when we consider the estimation of densities.

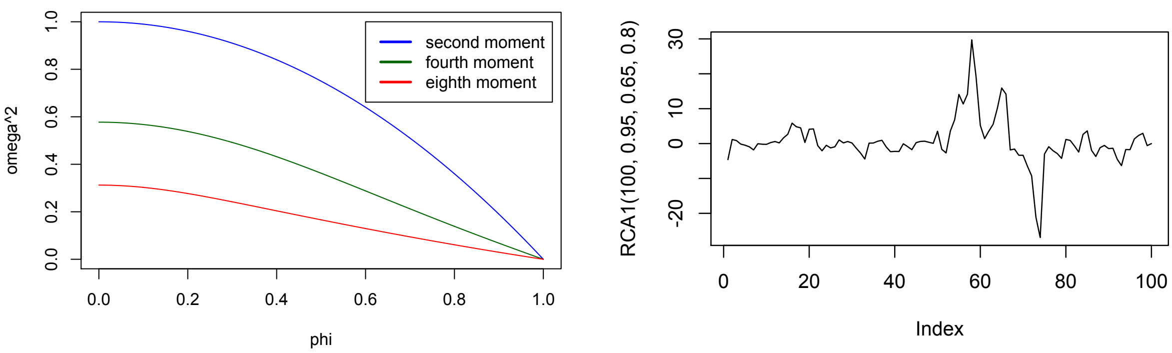

Figure 2.2: Permitted parameters for finite moments (left), sample path of a stationary RCA(1) process without finite moments (right)

In this and the following chapters, we will heavily make use of the so-called truncated RCA process that can be obtained from Equation (2.3) by just considering the last s terms of the innovation noise e_t and the disturbance noise b_t and that is given by the following Definition.

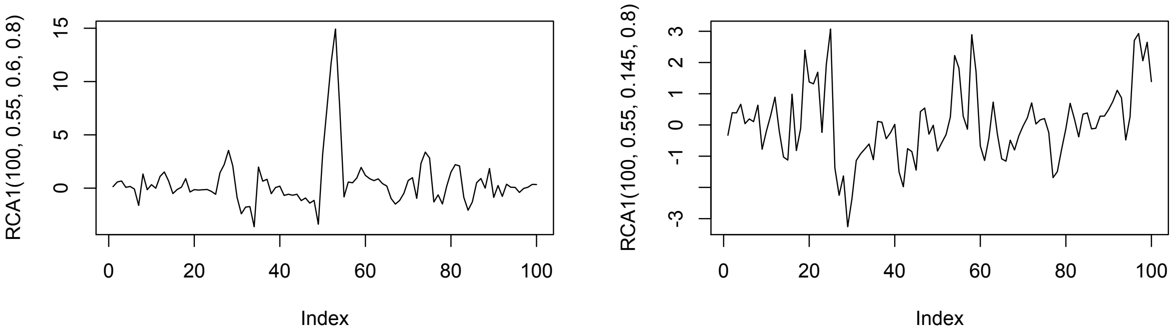

Figure 2.3: Sample paths of RCA(1) processes with finite second (left) and finite eight moments (right)

Definition 2.1.5. *In the context of Definition 2.1.1 the truncated version of an RCA process X_t is given by*

$$\tilde{X}_t^s = \sum_{i=0}^{s-1} \left(\prod_{j=0}^{i-1} (\varphi + b_{t-j}) \right) e_{t-i}. \tag{2.6}$$

Lemma 2.1.6. *If X_t has a finite second order moment, the L_2-norm of the difference between the original process X_t and the truncated process $\tilde{X}_t^s$ decreases to zero exponentially with increasing cut off s of $\tilde{X}_t^s$, i.e. with going further in the past:*

$$\boldsymbol{E}\left[\left(X_t - \tilde{X}_t^s \right)^2 \right] = \frac{\sigma^2}{1 - \varphi^2 - \omega^2} \left(\varphi^2 + \omega^2 \right)^s = C\vartheta^s, \ \vartheta < 1$$

The constants ϑ^s are absolutely summable.

Proof.

$$\mathbf{E}\left[\left(X_t - \tilde{X}_t^s \right)^2 \right] = \mathbf{E}\left[\left(\sum_{i=s}^{\infty} \prod_{j=0}^{i-1} (\varphi + b_{t-j}) e_{t-j} \right)^2 \right] = \frac{\sigma^2}{1 - \varphi^2 - \omega^2} \left(\varphi^2 + \omega^2 \right)^s := \vartheta^s$$

It directly follows that the constants ϑ^s are absolutely summable. □

Our intention in this and the next chapters is as follows: Having observations of X_t at hand, we would like to make inferences about the autorregressive parameter φ and about the second and the fourth moments of the innovation and the disturbance noise as well as about the distribution of the estimators for these paramters.
The problem is that we cannot directly observe b_t or e_t nor construct estimated residuals $\hat{b}_t$ or $\hat{e}_t$ after having estimated the parameter φ like in the standard autoregressive model. In the random coefficient autoregressive model, we can only construct convoluted residuals

$$\hat{u}_t = X_t - \hat{\varphi} X_{t-1} \stackrel{?}{=} \hat{b}_t X_{t-1} + \hat{e}_t$$

and cannot split up this sum to obtain individual residuals $\hat{b}_t$ and $\hat{e}_t$.
For simpler notation and computational ease we assume for the remainder that we have $n + 1$ observations $X_0, \ldots, X_n$ at hand and that not all observations are equal.

2.2 Parameter estimation

2.2.1 Existing estimators

Several ways to estimate the AR parameter φ as well as the variances of b_t and e_t, ω^2 and σ^2, respectively, can be found in the literature. Nicholls & Quinn (1980) and Nicholls & Quinn (1982) introduced least squares estimators, Nicholls & Quinn (1982) and Aue et al. (2006) developed an approach to estimate all three parameters simultaneously via a quasi maximum likelihood method, other estimators for φ were, amongst others, proposed by Schick (1996) and Koul & Schick (1996).
The approach of Nicholls & Quinn (1980) follows a three-step procedure: First, estimate φ via the well-known least squares estimator $\hat{\varphi} = \frac{\sum_{t=1}^{n} X_{t-1}X_t}{\sum_{t=1}^{n} X_{t-1}^2}$. Then, construct estimated residuals $\hat{u}_t = X_t - \hat{\varphi}X_{t-1}$ and estimate ω^2 by $\hat{\omega}^2 = \frac{\sum_{t=1}^{n} \hat{u}_t^2\left(X_{t-1}^2 - \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2\right)}{\sum_{t=1}^{n}\left(X_{t-1}^2 - \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2\right)^2}$. Finally, estimate σ^2 by $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2 - \hat{\omega}^2\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2$. They give the following result for the asymptotic distribution:

Theorem 2.2.1 (Nicholls & Quinn (1980)). *If X_t has finite fourth moments,*

$$\sqrt{n}\,(\hat{\varphi} - \varphi) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2}{\boldsymbol{E}[X_0^2]} + \omega^2 \frac{\boldsymbol{E}[X_0^4]}{\left(\boldsymbol{E}[X_0^2]\right)^2}\right).$$

If X_t has finite eight moments, $\sqrt{n}\,(\hat{\omega}^2 - \omega^2) \xrightarrow{D} \mathcal{N}\,(0, \tau^2)$ and $\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} \mathcal{N}\,(0, \nu^2)$ with τ^2, $\nu^2 > 0$, depending on the first eight moments of the process X_t as well as both the innovation and disturbance noise, see Nicholls & Quinn (1980). $\square$

To obtain confidence intervals for $\hat{\varphi}$, one has to estimate $\mathbf{E}\left[X_t^2\right]$ and $\mathbf{E}\left[X_t^4\right]$ as well as ω^2 and σ^2 from the data. We will make use of Theorem 2.2.1 in Chapter 3 to present a bootstrap approach to obtain confidence intervals and an approximation of the distribution of the estimator for finite sample sizes without any additional estimators needed.
The huge drawback of the least squares estimator is that there are put strong moment assumptions on the process to show consistency and that the parameter space is therefore very limited. A way to circumvent these is the quasi maximum likelihood estimator for the three parameters φ, ω^2, and σ^2 also introduced by Nicholls & Quinn (1982). Aue et al. (2006) considered the quasi maximum likelihood estimator in more detail and showed a central limit theorem under very mild regularity conditions. Except some regularity conditions on the sequences $(b_t, e_t)_{t\in\mathbb{Z}}$ and the stationarity conditions (2.2) on X_t, Aue et al. (2006) especially do not need any moment conditions on X_t, which allows for more flexibility in choosing the parameters. The approach is as follows: Suppose that b_t and e_t are normally distributed, then the conditional distribution of X_t given X_{t-1} is normal as well and for $u = (s, x, y)$ the log-likelihood-function and the QML-estimator

$$l_n\left((\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2)\right) = \inf_{u\in\Gamma} l_n(u) = \inf_{u\in\Gamma} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \ln(xX_{i-1}^2 + y)\right) \tag{2.7}$$

can be defined where $\Gamma = \left\{u = (s, x, y) : \ -s_0 \leqslant s \leqslant s_0,\ \frac{1}{x_0} \leqslant x \leqslant x_0,\ \frac{1}{y_0} \leqslant y \leqslant y_0\right\}$. For the following, we set further

$$g(X_i, u) = \frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \ln(xX_{i-1}^2 + y). \tag{2.8}$$

Aue et al. (2006) arrive at the following theorem:

Theorem 2.2.2. *If the inequalities (2.2) apply, and $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t\in\mathbb{Z}}$ are independent sequences of i.i.d. centered random variables with variance ω^2 and σ^2, respectively, and $\theta \in int\Gamma$, it holds for the estimator (2.7) that*

$$\sqrt{n}\left(\left(\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2\right)^\top - \left(\varphi, \omega^2, \sigma^2\right)^\top \right) \xrightarrow{D} \mathcal{N}\left(0, H^{-1}AH^{-1}\right),$$

where $A = \boldsymbol{E}\left[g'(X_1,\theta)^\top g'(X_1,\theta)\right]$, $H = \boldsymbol{E}\left[g''(X_1,\theta)\right]$. The independence of the sequences $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t\in\mathbb{Z}}$ can be relaxed, see Aue et al. (2006) for details.

For the proof we refer to Aue et al. (2006). □
A and H basically consist of terms of the form $\mathbf{E}\left[b_t^2\right]$, $\mathbf{E}\left[b_t^4\right]$, $\mathbf{E}\left[e_t^2\right]$, $\mathbf{E}\left[e_t^4\right]$, and

$$\mathbf{E}\left[\frac{X_{i-1}^{\kappa}}{\left(\omega^2 X_{i-1}^2 + \sigma^2\right)^{\gamma}}\right], \quad \gamma = 0, 1, 2, \quad \kappa = 0, \ldots, 2\gamma. \tag{2.9}$$

The problem when determining confidence intervals is the estimation of the moments of the white noises. It is not clear how to estimate the fourth moments without additional assumptions. In the next subsection we will show how to estimate these moments, however, to show consistency, again finiteness of the eighth moments of X_t is needed. Therefore, we will present a bootstrap approach to obtain these confidence intervals and an approximation of the distribution of the estimator in finite sample sizes without any additional moment assumptions in Chapter 4.

2.2.2 Estimating fourth order moments

In this subsection, we would like to introduce an estimator for the fourth order moments of an RCA process because we do need these estimators for our bootstrap procedure in Chapter 3. In doing so, we proceed similar to Nicholls & Quinn (1980) who estimated second order moments. We know that on the one hand $u_t = X_t - \varphi X_{t-1}$ and on the other hand $u_t = b_t X_{t-1} + e_t$, hence

$$\mathbf{E}\left[u_t^4\right] = \alpha^4 \mathbf{E}\left[X_{t-1}^4\right] + 6\omega^2\sigma^2 \mathbf{E}\left[X_{t-1}^2\right] + \beta^4.$$

Equipped with estimators $\hat{\varphi}$, $\hat{\sigma}^2$, $\hat{\omega}^2$ we can determine estimated residuals

$$\hat{u}_t = X_t - \hat{\varphi} X_{t-1}$$

and define shifted residuals

$$\tilde{u}_t = \hat{u}_t^4 - 6\hat{\omega}^2\hat{\sigma}^2 X_{t-1}^2$$

to obtain

$$\mathbf{E}\left[\tilde{u}_t\right] = \beta^4 + \alpha^4 \mathbf{E}\left[X_{t-1}^4\right],$$

and consequently (with $\epsilon_1, \dots \epsilon_n$ i.i.d. with mean 0),

$$\tilde{u}_t = \beta^4 + \alpha^4 X_{t-1}^4 + \epsilon_t,$$

so that we can finally regress $\tilde{u}_t$ on 1 and X_{t-1}^4 to receive the LS-estimators. We obtain

Theorem 2.2.3. *Assume that $\hat{\varphi}$, $\hat{\omega}^2$, $\hat{\sigma}^2$ are $\sqrt{n}$-consistent estimators for φ, ω^2, σ^2 and that the eight moment of X_t exists. Then, the fourth order moments $\alpha^4 = \boldsymbol{E}[b_t^4]$ and $\beta^4 = \boldsymbol{E}[e_t^4]$ can be estimated by the least squares estimators*

$$\begin{aligned}\hat{\alpha}^4 &= \frac{\sum_{t=1}^n \left(\left(\hat{u}_t^4 - 6\hat{\omega}^2\hat{\sigma}^2 X_{t-1}^2 \right) \left(X_{t-1}^4 - \frac{1}{n}\sum_{t=1}^n X_{t-1}^4 \right) \right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \frac{1}{n}\sum_{t=1}^n X_{t-1}^4 \right)^2} \\ \hat{\beta}^4 &= \frac{1}{n}\sum_{t=1}^n \left(\hat{u}_t^4 - 6\hat{\omega}^2\hat{\sigma}^2 X_{t-1}^2 \right) - \hat{\alpha}^4 \frac{1}{n}\sum_{t=1}^n X_{t-1}^4.\end{aligned} \tag{2.10}$$

Further, these estimators are consistent for α^4 and β^4, respectively.

Proof. We write $\overline{X^d} = \frac{1}{n}\sum_{t=1}^n X_t^d$, and replace $\hat{u}_t$, $\hat{\omega}^2$, $\hat{\sigma}^2$ in $\hat{\alpha}^4$ and $\hat{\beta}^4$ by the true parameters to define

$$\begin{aligned}\tilde{\alpha}^4 &= \frac{\sum_{t=1}^n \left(\left(u_t^4 - 6\omega^2\sigma^2 X_{t-1}^2 \right) \left(X_{t-1}^4 - \frac{1}{n}\sum_{t=1}^n X_{t-1}^4 \right) \right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \frac{1}{n}\sum_{t=1}^n X_{t-1}^4 \right)^2} \\ \tilde{\beta}^4 &= \frac{1}{n}\sum_{t=1}^n \left(u_t^4 - 6\omega^2\sigma^2 X_{t-1}^2 \right) - \tilde{\alpha}^4 \frac{1}{n}\sum_{t=1}^n X_{t-1}^4.\end{aligned} \tag{2.11}$$

Then,

$$\begin{aligned}\hat{u}_t^4 - u_t^4 &= \left(\hat{u}_t^2 - u_t^2 \right) \left(\hat{u}_t^2 + u_t^2 \right) \\ &= \left(\hat{u}_t - u_t \right) \left(\hat{u}_t - u_t + 2u_t \right) \left(\left(\hat{u}_t - u_t \right) \left(\hat{u}_t - u_t + 2u_t \right) + 2u_t^2 \right) \\ &= \left(\left(\varphi - \hat{\varphi} \right)^4 X_{t-1}^4 + 6 \left(\varphi - \hat{\varphi} \right)^2 X_{t-1}^2 u_t^2 + 4 \left(\varphi - \hat{\varphi} \right)^3 X_{t-1}^3 u_t + 4 \left(\varphi - \hat{\varphi} \right) X_{t-1} u_t^3 \right).\end{aligned}$$

Hence, it holds, if a $\sqrt{n}$-consistent estimator $\hat{\varphi}$ for φ is given, that

$$\frac{\frac{1}{n}\sum_{t=1}^n \left(\hat{u}_t^4 - u_t^4 \right) X_{t-1}^4}{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4} \right)^2} = \frac{\mathcal{O}_P\left(\frac{1}{\sqrt{\sqrt{n}}} \right)}{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4} \right)^2} = \mathcal{O}_P\left(\frac{1}{\sqrt{n}} \right)$$

since X_t has finite eight moments and $\mathbf{E}\left[|b_t|\right]$ and $\mathbf{E}\left[|e_t^3|\right]$ are finite as well. The denominator is of order $\mathcal{O}(1)$ and strictly positive if not all observations are equal. With this result we obtain from Equations (2.10) and (2.11) under the assumptions that $\hat{\omega}^2$ and $\hat{\sigma}^2$ are $\sqrt{n}$-consistent estimators as well:

$$\begin{aligned}\hat{\alpha}^4 - \tilde{\alpha}^4 &= \frac{1}{\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4} \right)^2} \left(\sum_{t=1}^n \left(\hat{u}_t^4 - u_t^4 \right) X_{t-1}^4 - 6 \left(\hat{\omega}^2\hat{\sigma}^2 - \omega^2\sigma^2 \right) \sum_{t=1}^n X_{t-1}^6 \right. \\ &\qquad \left. - \overline{X^4} \sum_{t=1}^n \left(\hat{u}_t 6^4 - u_t^4 \right) + 6 \left(\hat{\omega}^2\hat{\sigma}^2 - \omega^2\sigma^2 \right) \overline{X^4} \sum_{t=1}^n X_{t-1}^2 \right) \\ &= \mathcal{O}_P\left(\frac{1}{\sqrt{n}} \right).\end{aligned}$$

By using the definition of $\overline{X^d}$ and $\tilde{\alpha}^4$ and some algebra we obtain further:

$$
\begin{aligned}
&\tilde{\alpha}^4 - \alpha^4 \\
&= \tilde{\alpha}^4 - \alpha^4 \frac{\overline{X^8} - \overline{X^4}^2}{\overline{X^8} - \overline{X^4}^2} \\
&= \tilde{\alpha}^4 - \frac{\sum_{t=1}^n \beta^4 \left(X_{t-1}^4 - \overline{X^4}\right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)^2} - \alpha^4 \frac{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^8 - X_{t-1}^4 \overline{X^4}\right)}{\frac{1}{n}\sum X_{t-1}^8 + \frac{1}{n^2}\sum\sum X_{t-1}^4 X_{s-1}^4 - \frac{2}{n}\sum X_{t-1}^4 \frac{1}{n}\sum X_{t-1}^4} \\
&= \frac{\sum_{t=1}^n \left(\left(u_t^4 - 6\omega^2\sigma^2 X_{t-1}^2\right)\left(X_{t-1}^4 - \overline{X^4}\right)\right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \frac{1}{n}\overline{X^4}\right)^2} - \frac{\sum_{t=1}^n \beta^4 \left(X_{t-1}^4 - \overline{X^4}\right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)^2} \\
&\qquad - \frac{\sum_{t=1}^n \alpha^4 X_{t-1}^4 \left(X_{t-1}^4 - \overline{X^4}\right)}{\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)^2} \\
&= \frac{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)\left(u_t^4 - 6\omega^2\sigma^2 X_{t-1}^2 - \beta^4 - \alpha^4 X_{t-1}^4\right)}{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)^2} \\
&= \frac{1}{\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 - \overline{X^4}\right)^2} \left(\frac{1}{n}\sum_{t=1}^n \left(X_{t-1}^4 u_t^4 - 6\omega^2\sigma^2 X_{t-1}^6 - \beta^4 X_{t-1}^4 - \alpha^4 X_{t-1}^8\right) - \right. \\
&\qquad \left. \overline{X^4}\frac{1}{n}\sum_{t=1}^n \left(u_t^4 - 6\omega^2\sigma^2 X_{t-1}^2 - \beta^4 - \alpha^4 X_{t-1}^4\right)\right) \\
&= o_P(1)
\end{aligned}
$$

where we note that $\mathbf{E}\left[u_t^4\right] = \alpha^4 \mathbf{E}\left[X_{t-1}^4\right] + 6\sigma^4\omega^2 \mathbf{E}\left[X_{t-1}^2\right] + \beta^4$ and that X_t has finite eight moments and that therefore each sum in the expression above has expectation zero. This yields the assertion. Finally, the consistency of $\hat{\beta}^4$ can be obtained by similar computations. $\square$

2.3 A simulation study

For the two parameter sets that Nicholls and Quinn (1980) considered, the eighth moment of X_t does not exists, so we cannot use these sets for our estimators $\hat{\alpha}$ and $\hat{\beta}$. We slightly modify their first parameter set to arrive at our parameter set I with $(\varphi, \omega^2, \sigma^2) = (0.55, 0.145, 0.8)$ and this time both noises are exceptionally normally distributed. Thus, we can easily determine the fourth moments by $\alpha^4 = 3\omega^4$ and $\beta^4 = 3\sigma^4$ and the other moments by $\mathbf{E}\left[b_t^6\right] = 15\omega^6$ and $\mathbf{E}\left[b_t^8\right] = 105\omega^8$. Hence, $\varphi^8 + 28\varphi^6\omega^2 + 70\varphi^4\alpha^4 + 28\varphi^2\mathbf{E}\left[b_t^6\right] + \mathbf{E}\left[b_t^8\right] = 0.9589$ and from Equation (2.5) $\mathbf{E}\left[X_t^8\right] > 1037$. The parameters are chosen in such a way, that the value of Equation (2.4) is very close to one. That means that $\mathbf{E}\left[X_t^8\right]$ is very large and for a slight change in the parameters it does not exist anymore. The results of the simulation and estimation of all second and fourth moments are displayed in Figure 2.4 for sample size $n = 100$ (left) and $n = 10\,000$ (right). It becomes obvious that the estimators have a very high variability and that for small sample sizes they do not deliver the true values very exactly. For the large sample size the distribution of the estimators is close to the limiting distribtuion, however, the variability in the estimators is still high. All results are based on $M = 1\,000$ repetitions.

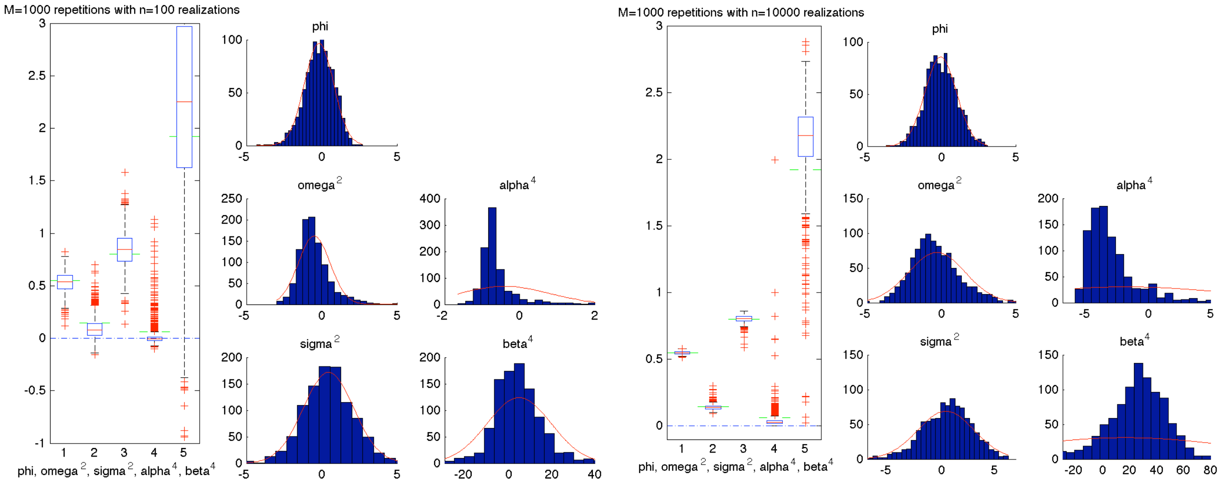

Figure 2.4: Estimators for $\hat{\varphi}$, $\hat{\omega}^2$, $\hat{\sigma}^2$, $\hat{\alpha}^4$ and $\hat{\beta}^4$, parameter set I, $n = 100$ (left) and $n = 10\,000$ (right): Boxplot and Histogram of centralized estimators stretched with $\sqrt{n}$.

To explain this behavior of the estimators we consider the problem of estimating the moments of X_t, because implicitly we use empirical moments in the estimators $\hat{\alpha}^4$ and $\hat{\beta}^4$ and refer to Figure 2.5 for the results. It becomes obvious that the variability of the estimator increases with increasing order of the moments under consideration and is tremendous for orders larger than two.

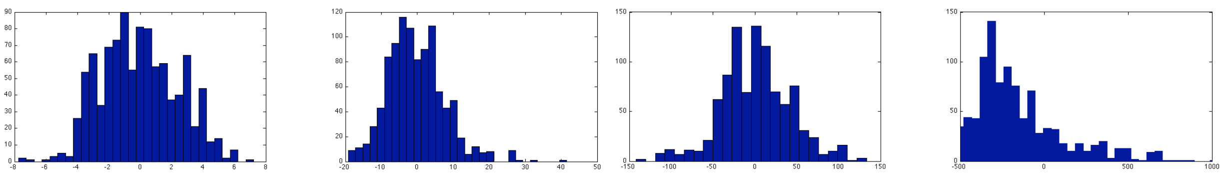

Figure 2.5: (Centralized) Estimators for $\mathbf{E}[X_t]$, $\mathbf{E}[X_t^2]$, $\mathbf{E}[X_t^3]$, and $\mathbf{E}[X_t^4]$, $n = 1000$, parameter set I

Another factor is that the eighth moment of X_t that we use to determine $\hat{\alpha}^4$ is extremely large since we chose parameters that are close to parameters for which the eighth moment does not exists anymore.
The behavior of the estimators is much better, especially for small sample sizes, for parameters that lead to a much smaller value for the eighth moment of X_t. To show this, we choose the parameters to be $(\varphi, \omega^2, \sigma^2) = (0.45, 0.1, 0.3)$. Again, both of the noises are normally distributed. This set is just for illustration purpose in this case and will not be considered further. Here, $\varphi^8 + 28\varphi^6\omega^2 + 210\varphi^4\omega^4 + 420\varphi^2\omega^6 + 105\omega^8 = 0.2066$. We refer to Figure 2.6 that shows satisfactory results for sample size $n = 100$ (left) and $n = 10\,000$ (right).

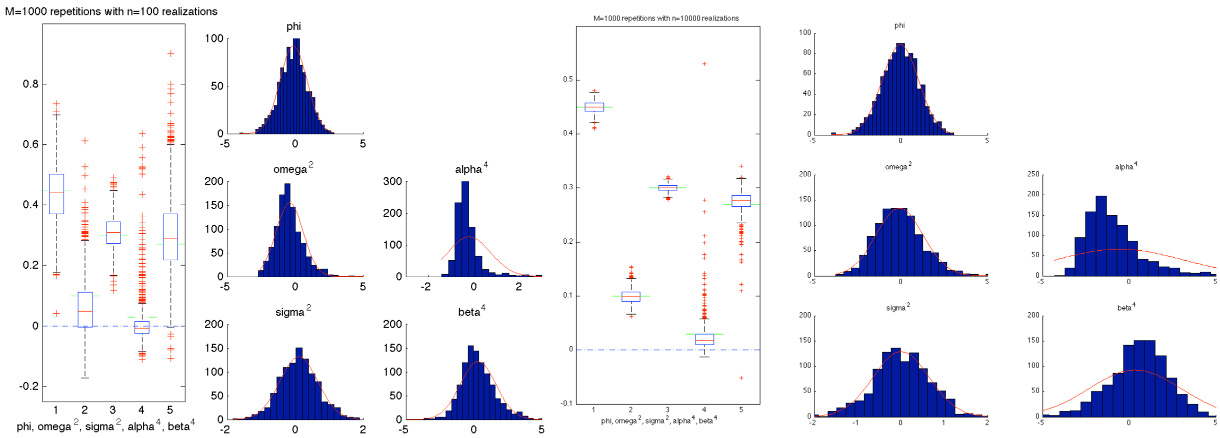

Figure 2.6: Estimators for $\hat{\varphi}$, $\hat{\omega}^2$, $\hat{\sigma}^2$, $\hat{\alpha}^4$ and $\hat{\beta}^4$, modified parameter set, $n = 100$ (left) and $n = 10\,000$ (right): Boxplot and Histogram of centralized estimators streched with $\sqrt{n}$.

For the parameter set II consistency cannot be established for all of the estimators we just considered. However, we can determine the QML estimator for these processes.
We simulate $n = 100$ realizations of the process X_t with e_t double exponentially distributed, b_t normally distributed and $(\varphi, \omega^2, \sigma^2) = (0.55,\ 0.6,\ 0.8)$. For an illustration we refer to Figure 2.7. The results are based on $M = 1\,000$ repetitions. It can be seen that for a small sample size the estimators are already very good, much better than the previously considered estimators. The variability is much smaller, especially for the estimators of the moments. Even though they are very close to the asymptotic distribution it is obvious that especially the last two mentioned estimators are skewed.
In the following chapters, we always assume that a strictly stationary solution to equation (2.1) that is given by equation (2.3) exists and that therefore the condition $\varphi^2 + \omega^2 < 1$ is fulfilled. In addition, we assume that we are provided with an $\sqrt{n}$-consistent estimator $\hat{\varphi}$ for the deterministic autoregressive parameter φ.

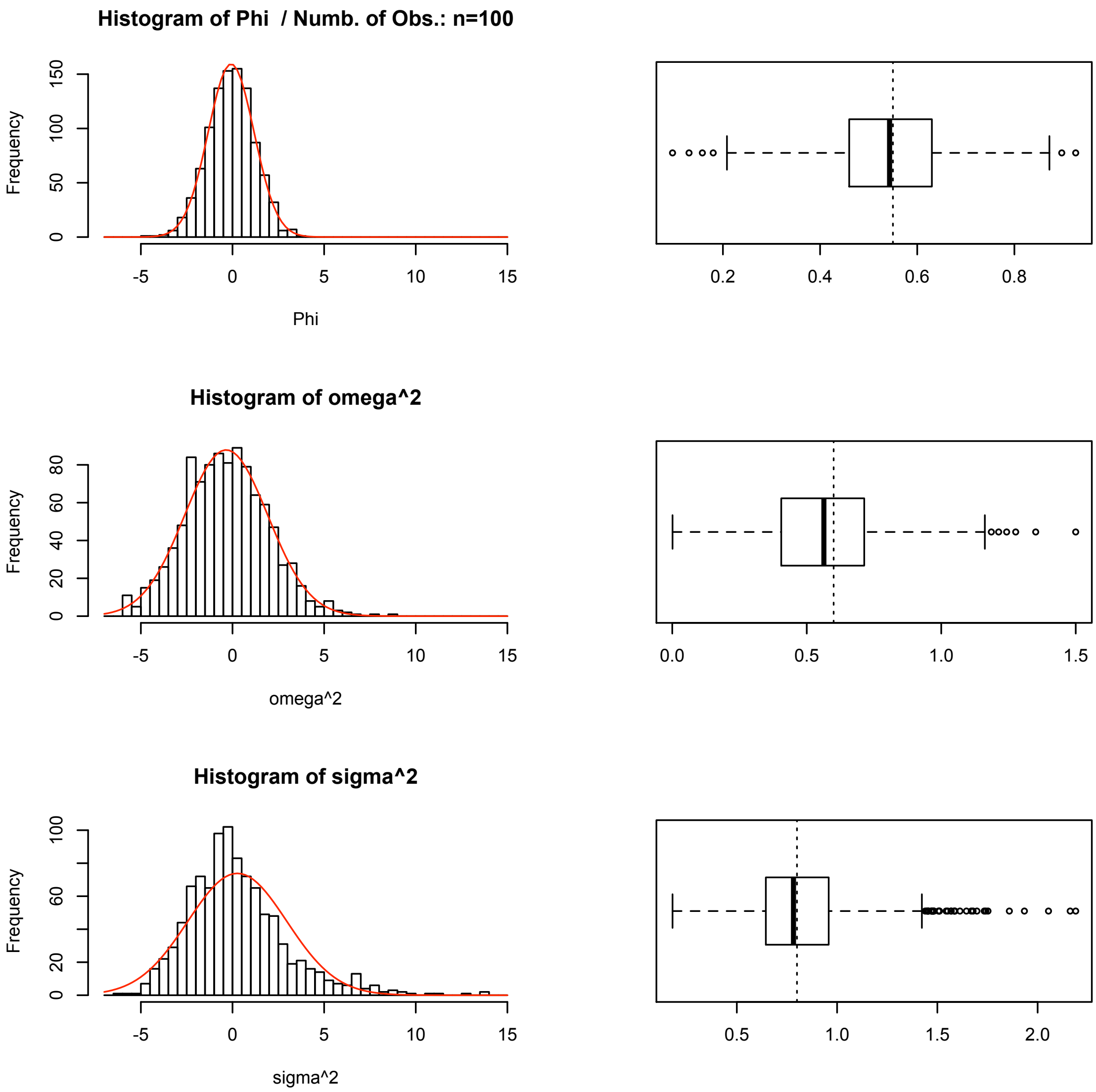

Figure 2.7: QML-estimator: Distribution of the parameters, $n = 100$

3 Bootstrap for the AR-parameter of random coefficient autoregressive processes

In this chapter, we first want to generalize existing bootstrap methods for standard autoregressive processes to RCA processes and thereafter further generalize these in the form of two wild bootstrap methods. These bootstrap methods are working for the distribution of the autoregressive parameter φ and are based on the least squares estimator given in Section 2.2.1. They also make use of the parameter estimator introduced in Section 2.2.2. Throughout this chapter, we assume that the process X_t has finite fourth moments and that the odd moments of the two noise sequences up to order seven are zero. In addition, we require to have consistent estimators for the fourth moments of both the noise sequences at hand. Using the estimators introduced in Chapter 2 this also means that the process X_t has to have finite eighth moments.

3.1 Existing bootstrap methods for AR(1) processes

Freedman (1984) and Efron & Tibshirani (1986) introduced a residual bootstrap for standard autoregressive processes. Kreiss (1988) and Kreiss (1997) and also Kreiss et al. (2011) reconsidered this residual based bootstrap and proved the validity of the autoregressive sieve bootstrap and a wild bootstrap procedure. The residual and the wild bootstrap work as follows: Having the estimator $\hat{\varphi}$, first construct estimated residuals $\hat{u}_t = X_t - \hat{\varphi} X_{t-1}$. Then, construct bootstrap realizations $u_1^*, \ldots, u_n^*$ i.i.d. on $\{\hat{u}_1, \ldots, \hat{u}_n\}$ for the standard residual based bootstrap and $u_i^w = \hat{u}_i \cdot K_i,\ K_i \sim (0,1)$ for the wild bootstrap. Finally, compute bootstrap observations of the process by

$$X_t^* = \hat{\varphi} X_{t-1}^* + u_t^* \qquad \text{or} \qquad X_t^w = \hat{\varphi} X_{t-1} + u_t^w,$$

respectively. The estimator used is the least squares estimator.

However, this method does not work for RCA models since the convoluted innovations u_t are heteroscedastic with conditional variance $\omega^2 X_{t-1}^2 + \sigma^2$, so that we propose two other approaches, the first of which is very similar to the residual based bootstrap just mentioned.

Praskova (2003) already considered the RCA model and proposed a wild bootstrap procedure that is very similar to the one proposed by Kreiss (1997) for AR processes. However, the huge drawback of this procedure is that this bootstrap procedure is not able to capture any dependencies within the realizations and therefore the bootstrap realizations do not form a process anymore. Furthermore, it only works in the RCA model of first order and even stronger moment conditions on the process than the ones used by Nicholls & Quinn (1980) are required to prove validity.

3.2 Residual based bootstrap

We would like to generalize the classical residual based bootstrap introduced by Kreiss (1988). We note that $\mathbf{E}\left[\varphi X_{t-1}|X_{t-1}\right] = \varphi X_{t-1}$ and $\mathbf{E}\left[b_t X_{t-1}|X_{t-1}\right] = 0$ and consider now just these realizations X_t with $|X_{t-1}| < \varepsilon = \varepsilon(n) \overset{n\to\infty}{\longrightarrow} 0$. Then, we have heuristically that

$$X_t = \varphi X_{t-1} + b_t X_{t-1} + e_t = \varphi X_{t-1} + (e_t + \delta_t) \approx \varphi X_{t-1} + e_t$$

and $\mathbf{E}\left[\delta_t|X_{t-1}\right] = 0$. This gives us approximative residuals for the innovation noise. Next, we consider just these realizations X_t with $|X_{t-1}| \geqslant M = M(n) \overset{n\to\infty}{\longrightarrow} \infty$. Note that, in particular, $X_{t-1} \neq 0$ in this case and that $\mathbf{E}\left[\frac{e_t}{X_{t-1}}\middle| X_{t-1}\right] = 0$. Then, we have heuristically that

$$\frac{X_t}{X_{t-1}} = \varphi\frac{X_{t-1}}{X_{t-1}} + b_t\frac{X_{t-1}}{X_{t-1}} + \frac{e_t}{X_{t-1}} = \varphi + b_t + \gamma_t \approx \varphi + b_t$$

and $\mathbf{E}\left[\gamma_t\right] = 0$. This gives us approximative residuals for the disturbance noise. The bootstrap procedure now is as follows.

Bootstrap Procedure 3.2.1. *Given $X_1, \ldots, X_n$ estimate φ by*

$$\hat{\varphi} = \frac{\sum_{t=1}^n X_{t-1}X_t}{\sum_{t=1}^n X_{t-1}^2}. \tag{3.1}$$

Denote by $2 \leqslant T_1 < T_2 < \cdots < T_{N_e}$, $N_e = \sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}$, the ordered random set of indices t for which $|X_{T_i-1}| < \varepsilon(n)$ is fulfilled and compute the related centered residuals

$$\hat{e}_{T_i} = X_{T_i} - \hat{\varphi}X_{T_i-1} - \frac{1}{N_e}\sum_{i=1}^{N_e}\left(X_{T_i} - \hat{\varphi}X_{T_i-1}\right).$$

Denote by $2 \leqslant S_1 < S_2 < \cdots < S_{N_b}$, $N_b = \sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}$, the ordered random set of indices t for which $|X_{S_i-1}| \geqslant M(n)$ is fulfilled and compute the related centered residuals

$$\hat{b}_{S_i} = \frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi} - \frac{1}{N_b}\sum_{i=1}^{N_b}\left(\frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi}\right).$$

This yields approximative residuals $\hat{e}_{T_1}, \ldots, \hat{e}_{T_{N_e}}$ and $\hat{b}_{S_1}, \ldots, \hat{b}_{S_{N_b}}$ with $N_e, N_b \leqslant n$. Generate independent bootstrap innovations $(e_t^*,\ t \in \mathbb{Z})$ by drawing with replacement from the set $\left\{\hat{e}_{T_1}, \ldots, \hat{e}_{T_{N_e}}\right\}$ and independent bootstrap disturbances $(b_t^*,\ t \in \mathbb{Z})$ by drawing with replacement from the set $\left\{\hat{b}_{S_1}, \ldots, \hat{b}_{S_{N_b}}\right\}$. This is the classical bootstrap idea.
Construct bootstrap observations of the process by

$$X_t^* = \left(\hat{\varphi} + b_t^*\right)X_{t-1}^* + e_t^*, \quad t \in \mathbb{Z}.$$

Asymptotical properties which $(M(n) : n \in \mathbb{N})$ and $(\varepsilon(n) : n \in \mathbb{N})$ have to fulfill can be found in the next subsection. Some practical remarks on how to choose the bandwidths can be found in Section 3.6. □

3.3 Validity of the residual based bootstrap

For the remainder of the chapter and for the next chapter as well, we use the following definition that is due to Lohse (1987), Definition 1.2, if we state that a sequence exhibits a certain characteristic in probability (abbreviated *i.P.*).

Definition 3.3.1. *A sequence of mappings $\{X_n,\ n \in \mathbb{N}\}$ on $(\Omega, \mathcal{A}, P)$ with values in $\mathcal{S}$ exhibits the characteristic $\mathcal{E}$ in P-probability, if $\forall\, \varepsilon > 0$, $\forall\, n \in \mathbb{N}$, $\exists\, \Omega_{n,\varepsilon} \in \mathcal{A}$ with $P(\Omega_{n,\varepsilon}) \geqslant 1-\varepsilon$, such that for $\{\omega_n,\ n \in \mathbb{N}\}$ with $\omega_n \in \Omega_{n,\varepsilon}$ holds: $\{X_n(\omega_n),\ n \in \mathbb{N}\}$ exhibits the characteristic $\mathcal{E}$ as sequence in $\mathcal{S}$.*

Remark 3.3.2. *The usual definition of convergence in probability is included in this definition: $\{X_n\}$ exhibits the characteristic "convergence to a" in probability if and only if $X_n \overset{n\to\infty}{\longrightarrow} a$ i.P. in the usual definition (Lohse (1987), Satz 1.3).*

Remark 3.3.3. *We require all estimators that we consider to be consistent. This means that all moment conditions we impose on the parameters, for example $\varphi^2 + \omega^2 < 1$, hold true in probability if we replace the parameters by their estimators. Furthermore, if, for example, the series $\left\{(\varphi^2 + \omega^2)^r,\ r \in \mathbb{N}\right\}$ is absoulutely summable, it can be easily seen that the series $\left\{(\hat{\varphi}_n^2 + \hat{\omega}_n^2)^r,\ r \in \mathbb{N}\right\}$ is absolutely summable in probability.*
However, in finite sample sizes, it could happen that the estimators do not fulfill these conditions. If this is the case, we could rescale them, since we know that they have to meet these conditions, hence, let us assume that the estimators are such that

$$\hat{\varphi}^2 + \hat{\omega}^2 < 1, \tag{3.2}$$

$$\hat{\varphi}^4 + 6\hat{\varphi}^2\hat{\omega}^2 + \hat{\alpha}^4 < \hat{\varphi}^2 + \hat{\omega}^2 < 1. \tag{3.3}$$

We recall that we denote the density of the stationary distribution of X_t by $f(\cdot)$ and the corresponding cumulative distribution function by $F(\cdot)$ and introduce the following conditions a selection of which will be assumed to hold for the process X_t in the following:

$$X_t \text{ has a density that is positive on the whole real axis and continuous in } 0. \tag{3.4}$$

$$X_t \text{ has a density that is positive and twice continuously differentiable in zero.} \tag{3.5}$$

$$X_t \text{ has finite absolute moments of order } 4+\vartheta \text{ for a } \vartheta > 0. \tag{3.6}$$

$$\text{The bandwidth } \varepsilon = \varepsilon(n) \overset{n\to\infty}{\longrightarrow} 0 \text{ is chosen such that } n^2\varepsilon^3 \overset{n\to\infty}{\longrightarrow} \infty. \tag{3.7}$$

$$M = M(n) \overset{n\to\infty}{\longrightarrow} \infty \text{ is such that } \sqrt{n}(1-F(M)) \overset{n\to\infty}{\longrightarrow} \infty \text{ and } \sqrt{n}F(-M) \overset{n\to\infty}{\longrightarrow} \infty. \tag{3.8}$$

Similar to Theorem 2.1.2 we have for the moments of the bootstrap process:

Theorem 3.3.4. *Under Conditions (4.2) through (4.5) it holds for the bootstrap process of Procedure 4.1.1*

$$\begin{aligned} \boldsymbol{E}^*\left[X_t^{*^2}\right] = \mathcal{O}_P(1) \quad &\Longleftrightarrow \quad \boldsymbol{E}\left[X_t^2\right] < \infty \\ \boldsymbol{E}^*\left[X_t^{*^4}\right] = \mathcal{O}_P(1) \quad &\Longleftrightarrow \quad \boldsymbol{E}\left[X_t^4\right] < \infty. \end{aligned}$$

Further, if $\boldsymbol{E}[X_t^2] < \infty$ and $\boldsymbol{E}[X_t^4] < \infty$, respectively,

$$\boldsymbol{E}^*\left[X_t^{*2}\right] = \frac{\hat{\sigma}^2}{1-\hat{\varphi}^2-\hat{\omega}^2} \xrightarrow{i.P.} \frac{\sigma^2}{1-\varphi^2-\omega^2} = \boldsymbol{E}\left[X_t^2\right]$$

and

$$\boldsymbol{E}^*\left[X_t^{*4}\right] = \frac{\hat{\beta}^4 + 6\hat{\sigma}^4\frac{\hat{\varphi}^2+\hat{\omega}^2}{1-\hat{\varphi}^2-\hat{\omega}^2}}{1-\hat{\varphi}^4-6\hat{\varphi}^2\hat{\omega}^2-\hat{\alpha}^4} \xrightarrow{i.P.} \frac{\beta^4 + 6\sigma^4\frac{\varphi^2+\omega^2}{1-\varphi^2-\omega^2}}{1-\varphi^4-6\varphi^2\omega^2-\alpha^4} = \boldsymbol{E}\left[X_t^4\right].$$

The proof can be found in Section 3.7. □

Theorem 3.3.5. *If Conditions (4.2) through (4.5) are valid, the following convergence holds true for the parameter estimate $\hat{\varphi}^*$ generated by the bootstrap procedure 4.1.1 and by replacing X_t by X_t^* in Equation (3.1):*

$$\sqrt{n}\,(\hat{\varphi}^* - \hat{\varphi}) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2}{\boldsymbol{E}[X_0^2]} + \omega^2\frac{\boldsymbol{E}[X_0^4]}{(\boldsymbol{E}[X_0^2])^2}\right) \; i.P.$$

The proof is delayed to Section 3.7. □

Remark 3.3.6. *For the Kolmogorov distance*

$$d_K\left(P_n, Q\right) = \sup_{x\in\mathbb{R}} |P_n\{(-\infty, x]\} - Q\{(-\infty, x]\}|$$

between two probability measures P_n and Q and Q having a continuous distribution function it holds true that

$$P_n \xrightarrow{D} Q \quad \Longleftrightarrow \quad d_K(P_n, Q) \xrightarrow{n\to\infty} 0$$

□

The consistency of the bootstrap procedure now follows directly:

Corollary 3.3.7. *Under the conditions of Theorem 3.3.5, the bootstrap procedure 4.1.1 is consistent for $\mathcal{L}\left(\sqrt{n}\,(\hat{\varphi}_n - \varphi)\right)$.*

Proof. Denote by $\xi^2 = \dfrac{\sigma^2}{\mathbf{E}[X_0^2]} + \omega^2\dfrac{\mathbf{E}[X_0^4]}{(\mathbf{E}[X_0^2])^2}$, by $\mathcal{L}_n$ the distribution of $\sqrt{n}\,(\hat{\varphi}_n - \varphi)$ and by $\mathcal{L}_n^*$ the one of $\sqrt{n}\,(\hat{\varphi}_n^* - \hat{\varphi}_n)$. Then, by Theorems 2.2.1 and 3.3.5,

$$d_K\left(\mathcal{L}_n, \mathcal{L}_n^*\right) \leqslant d_K\left(\mathcal{L}_n, \mathcal{N}\left(0, \xi^2\right)\right) + d_K\left(\mathcal{L}_n^*, \mathcal{N}\left(0, \xi^2\right)\right) \xrightarrow{n\to\infty} 0 \quad i.P..$$

□

3.4 Partially residual based bootstrap

Using the bootstrap method described above, it could happen that due to the distribution of X_t there are very few observations available that are large enough to be used to determine the disturbance residuals. However, if analyzing the asymptotic variance of the estimator $\hat{\varphi}$ that is given in Theorem 2.2.1, we can see that it only depends on the second and fourth moments of the process X_t and of the second moment of both the innovation and the disturbance noise. As we have seen in Theorem 2.1.2, the second and fourth moment of the process itself only depends on the second and fourth moments of both the innovation and the disturbance noise as well. Hence, it suffices to reproduce these moments correctly to obtain a bootstrap procedure that is consistent for φ.

Bootstrap Procedure 3.4.1. *Given $X_1, \ldots, X_n$ estimate φ by*

$$\hat{\varphi} = \frac{\sum_{t=1}^{n} X_{t-1} X_t}{\sum_{t=1}^{n} X_{t-1}^2}.$$

Denote by $2 \leqslant T_1 < T_2 < \cdots < T_N$, $N = \sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}$, the ordered random set of indices t for which $|X_{t-1}| < \varepsilon(n)$ is fulfilled and compute the related centered residuals

$$\hat{e}_{T_i} = X_{T_i} - \hat{\varphi} X_{T_i - 1} - \frac{1}{N} \sum_{i=1}^{N} \left(X_{T_i} - \hat{\varphi} X_{T_i-1} \right).$$

This yields approximative residuals $\hat{e}_{T_1}, \ldots, \hat{e}_{T_N}$ with $N \leqslant n$. Construct independent bootstrap innovations $(e_t^*,\ t \in \mathbb{Z})$ by sampling with replacement from the set $\{\hat{e}_{T_1}, \ldots, \hat{e}_{T_N}\}$. Given consistent estimators $\hat{\omega}^2$ and $\hat{\alpha}^4$ generate bootstrap realizations $(b_t^*,\ t \in \mathbb{Z})$ i.i.d. and independent from the $(e_t^*,\ t \in \mathbb{Z})$ with

$$\left(\boldsymbol{E}^* [b_t^*], \boldsymbol{E}^* [b_t^{*2}], \boldsymbol{E}^* [b_t^{*3}], \boldsymbol{E}^* [b_t^{*4}] \right) = \left(0, \hat{\omega}^2, 0, \hat{\alpha}^4 \right).$$

Construct bootstrap observations of the process by

$$X_t^* = \left(\hat{\varphi} + b_t^* \right) X_{t-1}^* + e_t^*, \quad t \in \mathbb{Z}.$$

The random variables b_t^* can be generated by any user defined method obeying the moment conditions stated above, for example a four-point distribution can be used. □

Theorem 3.4.2. *If Conditions (3.5), (3.6), and (4.4) are valid, the bootstrap procedure 3.4.1 is consistent for $\mathcal{L}\left(\sqrt{n}\left(\hat{\varphi}_n - \varphi\right)\right)$.*

Proof. The argumentation basically is the same as in the proof of Theorem 3.3.5 and Corollary 4.2.7 and is therefore omitted. □

3.5 Exclusively moment based bootstrap

Another variation is the following method that uses only the estimated moments of the both the innovation and disturbance parameters. As we have argued before, the asymptotic distribution of the estimator only depends on the second and fourth moments of both the innovation and the disturbance noise. Hence, the aforementioned method can be generalized so that we do not need to determine any residuals at all. It suffices to reproduce the moments of both the innovation and the disturbance noise correctly to obtain a bootstrap procedure that is consistent for φ :

Bootstrap Procedure 3.5.1. *Given estimators $\hat{\sigma}^2$ and $\hat{\omega}^2$ as well as $\hat{\beta}^4$ and $\hat{\alpha}^4$, generate bootstrap observations $(e_t^*,\ t \in \mathbb{Z})$ i.i.d. with*

$$\left(\boldsymbol{E}^*[e_t^*], \boldsymbol{E}^*[e_t^{*2}], \boldsymbol{E}^*[e_t^{*3}], \boldsymbol{E}^*[e_t^{*4}]\right) = \left(0, \hat{\sigma}^2, 0, \hat{\beta}^4\right)$$

and $(b_t^*,\ t \in \mathbb{Z})$ i.i.d. and independent from the $(e_t^*,\ t \in \mathbb{Z})$ with

$$\left(\boldsymbol{E}^*[b_t^*], \boldsymbol{E}^*[b_t^{*2}], \boldsymbol{E}^*[b_t^{*3}], \boldsymbol{E}^*[b_t^{*4}]\right) = \left(0, \hat{\omega}^2, 0, \hat{\alpha}^4\right).$$

Construct bootstrap observations of the process by

$$X_t^* = (\hat{\varphi} + b_t^*)\, X_{t-1}^* + e_t^*, \quad t \in \mathbb{Z}.$$

The random variables e_t^* and b_t^* can be generated by any user defined method, for example a four-point distribution can be used. □

Theorem 3.5.2. *Under Condition (3.6) the bootstrap procedure 3.5.1 is consistent for φ.*

Proof. The argumentation basically is the same as in the proof of Theorem 3.3.5 and Corollary 4.2.7 and is therefore omitted. □

We can also generalize the methods described above:

Remark 3.5.3. *In principle, the approach of Theorem 2.2.3 in estimating the fourth moments (see Nicholls & Quinn (1980)) and of the three aforementioned bootstrap approaches can be extended to a general random coefficient autoregressive model of order p,*

$$X_t = \sum_{i=1}^{p} (\varphi_i + b_{t,i})\, X_{t-i} + e_t. \tag{3.9}$$

For the moment based bootstrap, we just generate the replications of all the random variables as described above and plug them into the bootstrap analogon of Equation (3.9). For the residual based bootstrap, a combination of the last $p+1$ observations $X_t, \ldots, X_{t-p}$ being smaller than ε or larger than M, respectively, can be found to extract a certain residual $\hat{b}_{t-i}$ or $\hat{e}_{t-i}$ with the aforementioned procedures. These can again be used to obtain bootstrap replicates. However, it should be noted that these would be asymptotic results and that even for a process of order two, already a large sample size has to be available to obtain reliable results.

3.6 A simulation study

We chosse the parameter to be parameter set I, namely $(\varphi, \omega^2, \sigma^2) = (0.55, 0.145, 0.8)$, c.f. Chapter 2, and the disturbance noise b_t normally distributed and the innovation noise e_t double-exponentially distributed. We estimate the AR-parameter and the second moments with the least-squares method introduced by Nicholls & Quinn (1980) (see Section 2.2.1) and the fourth moments by the method given in Section 2.2.2.
We consider the distribution of the difference between the estimator and the true parameter $\mathcal{L}\left(\sqrt{n}\,(\hat{\varphi} - \varphi)\right)$ and determine the quantiles of the distribution at different levels. In doing so, we simulate n (e.g. $n = 50$ or $n = 100$) realizations of the process X_t first

and determine $\hat{\varphi}$ out of realizations $X_0, \ldots, X_n$. On the one hand we determine all parameters necessary for the limiting normal distribution and use the approximation via this normal distribution. On the other hand, we generate N (e.g. $N = 1000$) bootstrap processes of length n each to determine the bootstrap distribution $\mathcal{L}^* \left(\sqrt{n}\,(\hat{\varphi}^* - \hat{\varphi})\right)$ and out of this its desired quantiles. This is done with each bootstrap method described above. The parameters for the residual based bootstrap are chosen as $\varepsilon(n) = 2.3 \cdot n^{-\frac{1}{6}}$ and $M(n) = 1.4 \ln(\ln(n))$. It turned out in simulations before, that the boostrap method is not very sensitive to the choice of these bandwidths. Depending on the concrete sample, the number of realizations that are considered by these bandwidths varies very much.

All of this is repeated T times (e.g. $T = 1000$) to obtain the boxplots displayed in Figures 3.1 and 3.2: The dashed lines give the true (though simulated) quantiles, the dotted lines the quantiles of the limiting normal distribution with the true parameters. The blue boxplots on the left in the left plots give the quantiles of the limiting normal distribution with estimated parameters, and the red boxplots give the quantiles determined via the bootstrap approximations, from left to right these are the residual based bootstrap, the partially residual based bootstrap and the moment based bootstrap. These are also displayed in the middle and right plots (in same order, lower an upper quantile).

One can easily see that the bootstrap approximations perform considerably better than the approximation via the normal distribution. The normal approximation is not only much more spreading than the boostrap approximations, but most of the time it is also overestimating the distance of the quantile from zero. The median of the bootstrap approximations, however, nearly hits the true quantile. The middle plot and the plot on the right in each figure gives the results for the different bootstrap methods in the same order as before to compare the different approaches. One plot gives the lower quantile, one the upper one. It can be seen that altough the residual based bootstrap performs quite well, it is outperformed by the other two methods and that the moment based bootstrap yields the best results in this situation.

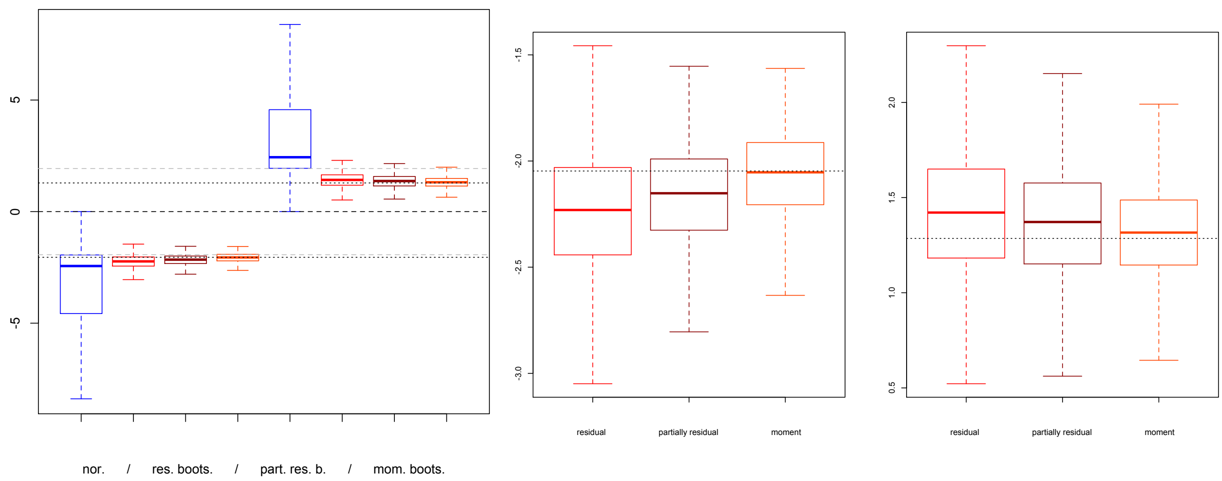

Figure 3.1: Quantiles: Bootstrap and normal approximation, $n = 50$, $\alpha = 10\%$, 90%

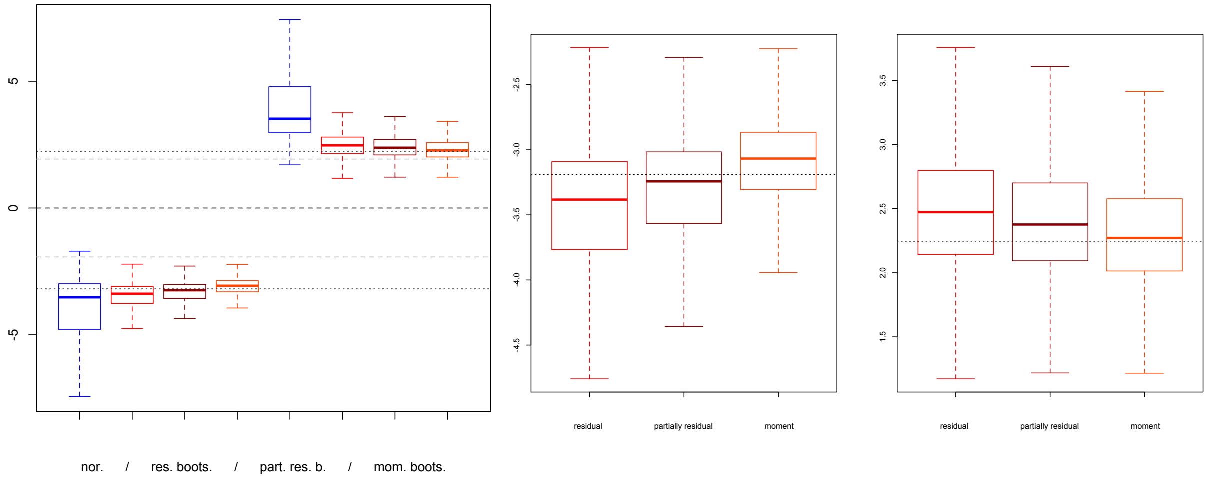

Figure 3.2: Quantiles: Bootstrap and normal approximation, $n = 100$, $\alpha = 1\%, 99\%$

3.7 Proofs

To show that the residual based bootstrap procedure works, we have to state some Lemmas and give a few Definitions first.
Before we consider the moments of the bootstrap variables, we state a Lemma first.

Lemma 3.7.1. *If Conditions (3.5) and (4.4) are valid, it holds true that*

$$\begin{aligned}
\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} &= f(0) + o_P(1),\\
\frac{1}{2n\varepsilon}\sum_{t=1}^{n}X_{t-1}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} &= o_P(1),\\
\frac{1}{2n\varepsilon}\sum_{t=1}^{n}X_{t-1}^2\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} &= o_P(1),\\
\frac{1}{2n\varepsilon}\sum_{t=1}^{n}b_t^q e_t^{(1-q)}X_{t-1}^d\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} &= \mathcal{O}_P\left(\frac{1}{\sqrt{n\varepsilon}}\right), \quad d=1,2,3,\ q=0,1,\\
\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|\geqslant M\}} &= F(-M)+F(M)+\mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right),\\
\frac{1}{n}\sum_{t=1}^{n}\frac{\mathbb{1}_{\{|X_{t-1}|\geqslant M\}}}{X_{t-1}^2} &= \frac{1}{M^2}\left(F(-M)+F(M)\right)+o(1).
\end{aligned}$$

Proof. The expectations of the terms can be determined as follows with partial integration and an $\tilde{\varepsilon} \in (-\varepsilon, \varepsilon)$:

$$\begin{aligned}
\frac{1}{2\varepsilon}\mathbf{E}\left[\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] &= \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}f(x)\mathrm{d}x = \frac{F(\varepsilon)-F(-\varepsilon)}{2\varepsilon}\\
&= \frac{2\varepsilon f(0) + c\varepsilon^3 f(\tilde{\varepsilon})}{2\varepsilon} = f(0) + \mathcal{O}\left(\varepsilon^2\right) \qquad (3.10)\\
\frac{1}{2\varepsilon}\mathbf{E}\left[X_{t-1}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] &= \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}xf(x)\mathrm{d}x = \frac{\varepsilon F(\varepsilon)+\varepsilon F(-\varepsilon)}{2\varepsilon} - \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}F(x)\mathrm{d}x
\end{aligned}$$

$$= \varepsilon f(0) - F(\tilde{\varepsilon}) + \mathcal{O}\left(\varepsilon^2\right) = \mathcal{O}(\varepsilon) \tag{3.11}$$

$$\frac{1}{2\varepsilon}\mathbf{E}\left[X_{t-1}^2 \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = \varepsilon\frac{F(\varepsilon)+F(-\varepsilon)}{2} - \tilde{\varepsilon}F(\tilde{\varepsilon}) + \mathcal{O}\left(\varepsilon^2\right) \tag{3.12}$$

$$\leqslant \varepsilon\frac{F(\varepsilon)+F(-\varepsilon)}{2} + \mathcal{O}\left(\varepsilon^2\right) = \varepsilon^2 f(0) + \mathcal{O}\left(\varepsilon^2\right) = \mathcal{O}\left(\varepsilon^2\right)$$

$$\frac{1}{2\varepsilon}\mathbf{E}\left[b_t^q e_t^{1-q} X_{t-1}^d \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = \mathbf{E}\left[b_t^q\right]\mathbf{E}\left[e_t^{1-q}\right]\frac{1}{2\varepsilon}\mathbf{E}\left[\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = 0 \tag{3.13}$$

since either $\mathbf{E}[b_t] = 0$ or $\mathbf{E}[e_t] = 0$. For the variance of the first expression we have

$$\begin{aligned}
&\mathbf{Var}\left[\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] \\
&= \frac{1}{4n^2\varepsilon^2}\sum_{t=1}^{n}\mathbf{Var}\left[\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] + \frac{1}{2n^2\varepsilon^2}\sum_{t=2}^{n}(n-i+1)\mathbf{Cov}\left[\mathbb{1}_{\{|X_0|<\varepsilon\}}, \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] \\
&\leqslant \frac{1}{4n\varepsilon^2}P(|X_0|<\varepsilon)(1-P(|X_0|<\varepsilon)) + \frac{1}{2n\varepsilon^2}\sum_{t=2}^{n}\mathbf{Cov}\left[\mathbb{1}_{\{|X_0|<\varepsilon\}}, \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} - \mathbb{1}_{\{|\tilde{X}_{t-1}^{t-1}|<\varepsilon\}}\right] \\
&\leqslant \frac{1}{2n\varepsilon}\left(f(0)+\mathcal{O}(\varepsilon^2)\right) + \frac{1}{2n\varepsilon^2}\sum_{t=2}^{n}\mathbf{E}\left[\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right]^{\frac{1}{2}}\mathbf{E}\left[\left(\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} - \mathbb{1}_{\{|\tilde{X}_{t-1}^{t-1}|<\varepsilon\}}\right)^2\right]^{\frac{1}{2}} \\
&\leqslant \frac{1}{2n\varepsilon}\left(f(0)+\mathcal{O}(\varepsilon^2)\right) + \frac{1}{\sqrt{2}n\varepsilon^{1.5}}\sqrt{\left(f(0)+\mathcal{O}\left(\varepsilon^2\right)\right)} \\
&\qquad \cdot\sum_{t=2}^{n}\left(P\left\{|X_{t-1}|<\varepsilon,\ |\tilde{X}_{t-1}^{t-1}|\geqslant\varepsilon\right\} + P\left\{|X_{t-1}|\geqslant\varepsilon,\ |\tilde{X}_{t-1}^{t-1}|<\varepsilon\right\}\right)^{\frac{1}{2}} \\
&\leqslant \frac{1}{2n\varepsilon}\left(f(0)+\mathcal{O}(\varepsilon^2)\right) + \frac{1}{\sqrt{2}n\varepsilon^{1.5}\delta^2}\sqrt{\left(f(0)+\mathcal{O}\left(\varepsilon^2\right)\right)}C\sum_{t=0}^{n}\vartheta^t \\
&\leqslant \frac{1}{2n\varepsilon}\left(f(0)+\mathcal{O}(\varepsilon^2)\right) + \frac{1}{\sqrt{2}n\varepsilon^{1.5}\delta^2}\sqrt{\left(f(0)+\mathcal{O}\left(\varepsilon^2\right)\right)}C\frac{1-\vartheta^{n-1}}{1-\vartheta} \\
&= o(1)
\end{aligned}$$

with $\vartheta < 1$ and $C \in \mathbb{R}$ and where we have used that

$$\begin{aligned}
P\left\{|X_{t-1}|<\varepsilon, |\tilde{X}_{t-1}^{t-1}|\geqslant\varepsilon\right\} &= P\left\{|\tilde{X}_{t-1}^{t-1}| - |X_{t-1}| \geqslant \varepsilon - |X_{t-1}| > 0,\ |X_{t-1}|<\varepsilon\right\} \\
&\leqslant P\left\{|\tilde{X}_{t-1}^{t-1} - X_{t-1}| > \delta\right\} \quad \text{for a } 0<\delta<\varepsilon \\
&\leqslant \frac{1}{\delta^2}\mathbf{E}\left[\left(\tilde{X}_{t-1}^{t-1} - X_{t-1}\right)^2\right] \\
&\leqslant \frac{\vartheta^t}{\delta^2}C', \quad \text{with } \tilde{\vartheta}<1,\ C'\in\mathbb{R},\ \text{and } \delta<\varepsilon.
\end{aligned}$$

With the assertions about the means (Equations (3.10) through (3.13)) and the inequality

$$\begin{aligned}
&\mathbf{E}\left[\left(X_{t-1}^d \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} - \tilde{X}_{t-1}^{d_{t-1}}\mathbb{1}_{\{|\tilde{X}_{t-1}^{t-1}|<\varepsilon\}}\right)^2\right] \\
&\qquad\qquad \leqslant \mathbf{E}\left[\left(X_{t-1}^d - \tilde{X}_{t-1}^{d_{t-1}}\right)^2\right] + \mathbf{E}\left[\left(\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} - \mathbb{1}_{\{|\tilde{X}_{t-1}^{t-1}|<\varepsilon\}}\right)^2\right],
\end{aligned}$$

which holds for $\varepsilon < 1$ since then $X_{t-1}^d \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} < \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}$, the variances of the second and the third term, respectively, can be limited by similar computations. For the fourth term consider

$$\frac{1}{4n^2\varepsilon^2}\sum_{t=1}^{n}\mathbf{Var}\left[e_t\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = \frac{1}{4n\varepsilon^2}\mathbf{E}\left[e_1^2\mathbb{1}_{\{|X_0|<\varepsilon\}}\right] = \mathbf{E}\left[e_1^2\right]\frac{1}{4n\varepsilon^2}\mathbf{E}\left[\mathbb{1}_{\{|X_0|<\varepsilon\}}\right] = \mathcal{O}\left(\frac{1}{n\varepsilon}\right)$$

and without loss of generality for $t < s$

$$\begin{aligned}\mathbf{Cov}\left[e_t\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}, e_s\mathbb{1}_{\{|X_{s-1}|<\varepsilon\}}\right] &= \mathbf{E}\left[e_t\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}e_s\mathbb{1}_{\{|X_{s-1}|<\varepsilon\}}\right]\\ &= \mathbf{E}\left[e_t\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\mathbb{1}_{\{|X_{s-1}|<\varepsilon\}}\right]\mathbf{E}\left[e_s\right] = 0\end{aligned}$$

For b_t and other powers of X_t the argumentation is exactly the same. The fifth term follows by a standard argumentation as the one that was used for the first term, and the last term as well, where it can be used that $\frac{\mathbb{1}_{\{|X_{t-1}|\geqslant M\}}}{X_{t-1}^2} \leqslant \frac{1}{M^2}\mathbb{1}_{\{|X_{t-1}|\geqslant M\}}$. □

With these results, we can evaluate the moments of e_t^* and b_t^*. Recall that the bootstrap realizations e_t^* are drawn independently with replacement from the random set of the approximative residuals $\left\{\hat{e}_{T_1}, \ldots, \hat{e}_{T_{N_e}}\right\}$ constructed by

$$\begin{aligned}\hat{e}_{T_i} &= X_{T_i} - \hat{\varphi}X_{T_i-1} - \frac{1}{N_e}\sum_{i=1}^{N_e}\left(X_{T_i} - \hat{\varphi}X_{T_i-1}\right)\\ &= X_{T_i} - \hat{\varphi}X_{T_i-1} - \frac{\sum_{t=1}^{n}\left(X_t - \hat{\varphi}X_{t-1}\right)\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}{\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}\end{aligned} \tag{3.14}$$

if $|X_{T_i-1}| < \varepsilon(n)$, so that we can state:

Lemma 3.7.2. *If Condition (4.4) is valid, it holds holds true that*

$$\boldsymbol{E}^*\left[e_t^*\right] = 0 = \boldsymbol{E}\left[e_t\right]$$

and

$$\boldsymbol{E}^*\left[{e_t^*}^2\right] \xrightarrow{i.P.} \sigma^2 = \boldsymbol{E}\left[e_t^2\right]$$

as well as

$$\boldsymbol{E}^*\left[{e_t^*}^4\right] \xrightarrow{i.P.} \beta^4 = \boldsymbol{E}\left[e_t^4\right]$$

Proof. The first equation follows immediately by construction of e_t^*. The expectation of the second term can with Equation (3.14) be determined as

$$\begin{aligned}\mathbf{E}^*\left[{e_t^*}^2\right] &= \frac{1}{N_e}\sum_{i=1}^{N_e}\hat{e}_{T_i}^2\\ &= \frac{1}{\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}\sum_{t=1}^{n}\hat{e}_t^2\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\\ &= \frac{1}{\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}\frac{1}{2n\varepsilon}\sum_{t=1}^{n}e_t^2\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\\ &\qquad + \frac{1}{\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\left(\hat{e}_t^2 - e_t^2\right)\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\\ &= \sigma^2 + o_P(1),\end{aligned}$$

what can be seen as follows: For the first summand combine

$$\mathbf{E}\left[\frac{1}{2n\varepsilon}\sum_{t=1}^{n} e_t^2 \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = \frac{1}{2n\varepsilon}\mathbf{E}[e_1^2]\sum_{t=1}^{n}\mathbf{E}[\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}] = \sigma^2 f(0) + o(1),$$
$$\mathbf{Var}\left[\frac{1}{2n\varepsilon}\sum_{t=1}^{n} e_t^2 \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right] = o(1), \quad \text{by similar computations as in Lemma 3.7.1,}$$

and

$$\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} = f(0) + o_P(1) \quad \text{(Lemma 3.7.1),}$$

and for the second summand consider

$$\begin{aligned}
&\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\left(\hat{e}_t^2 - e_t^2\right)\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} \\
&\quad = \left(\hat{\varphi}^2 - \varphi^2\right)\frac{1}{2n\varepsilon}\sum_{t=1}^{n} X_{t-1}^2 \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} + 2\left(\varphi - \hat{\varphi}\right)\frac{1}{2n\varepsilon}\sum_{t=1}^{n} X_t X_{t-1}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} \\
&\qquad + 2\frac{1}{2n\varepsilon}\sum_{t=1}^{n} b_t e_t X_{t-1}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} + \frac{1}{2n\varepsilon}\sum_{t=1}^{n} b_t X_{t-1}^2 \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} \\
&\qquad + \bar{e}^2\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} + 2\bar{e}\left(\hat{\varphi}\frac{1}{2n\varepsilon}\sum_{t=1}^{n} X_{t-1}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} - \frac{1}{2n\varepsilon}\sum_{t=1}^{n} X_t \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right) \\
&\quad = o_P(1)
\end{aligned}$$

where we have used that $\hat{\varphi}$ is a $\sqrt{n}$-consistent estimator for φ and Lemma 3.7.1 as well as the following result:

$$\begin{aligned}
\bar{e} &= \frac{1}{\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}}\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\left(X_t - \hat{\varphi}X_{t-1}\right)\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} \\
&= \frac{1}{\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}} \\
&\qquad \cdot\left(\frac{1}{2n\varepsilon}\sum_{t=1}^{n}\left[\varphi X_{t-1} + b_t X_{t-1} + e_t - \left(\varphi + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)\right)X_{t-1}\right]\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}\right) \\
&= o_P(1)
\end{aligned}$$

where we have used Lemma 3.7.1 again. The assertion about the fourth order moment follows by a similar argumentation. □

Recalling that the bootstrap realizations b_t^* are drawn independently with replacement from the random set $\{\hat{b}_{S_1}, \ldots, \hat{b}_{S_{N_b}}\}$ of the approximative residuals constructed by

$$\hat{b}_{S_i} = \frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi} - \frac{1}{N_b}\sum_{i=1}^{N_b}\left(\frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi}\right) = \frac{X_{t_i}}{X_{t_i-1}} - \hat{\varphi} - \frac{\sum_{t=1}^{n}\left(\frac{X_t}{X_{t-1}} - \hat{\varphi}\right)\mathbb{1}_{\{|X_{t-1}|\geqslant M\}}}{\sum_{t=1}^{n}\mathbb{1}_{\{|X_{t-1}|\geqslant M\}}} \tag{3.15}$$

if $|X_{S_i-1}| \geqslant M$, we can state in a similar manner to the previous Lemma:

Lemma 3.7.3. *Under Condition (4.5) it holds true that*

$$\boldsymbol{E}^{*}\left[b_t^*\right] = 0 = \boldsymbol{E}\left[b_t\right],$$
$$\boldsymbol{E}^{*}\left[b_t^{*2}\right] \xrightarrow{i.P.} \omega^2 = \boldsymbol{E}\left[b_t^2\right],$$
$$\boldsymbol{E}^{*}\left[b_t^{*4}\right] \xrightarrow{i.P.} \alpha^4 = \boldsymbol{E}\left[b_t^4\right].$$

Proof. The first equation follows immediately and for the other equation we follow the argumentation of Lemma 3.7.2. We define $\bar{b}$ analogously to $\bar{e}$ and obtain $\bar{b} = o_P(1)$. Further, the term

$$\frac{F(-M)}{F(-M) + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)} = 1 + \frac{F(-M)}{F(-M) + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)} - \frac{F(-M)}{F(-M)} = 1 + \frac{\mathcal{O}_P(1)}{\sqrt{n}F(-M) + \mathcal{O}_P\left(\frac{1}{n}\right)}$$

vanishes asymptotically if and only if $\sqrt{n}F(-M) \xrightarrow{n\to\infty} \infty$. Using this argumentation, Lemma 3.7.1, the fact that $\hat{\varphi}$ is $\sqrt{n}$-consistent for φ, and finally the argumentation of Lemma 3.7.2 we obtain from Equation (3.15):

$$\begin{aligned}
\mathbf{E}^{*}\left[b_t^{*2}\right] &= \frac{1}{N_b}\sum_{i=1}^{N_b} \hat{b}_{S_i}^2 \\
&= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}} \sum_{t=1}^n \hat{b}_t^2 \mathbb{1}_{\{|X_{t-1}|\geqslant M\}} \\
&= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}} \sum_{t=1}^n \Bigg((\varphi-\hat{\varphi})^2 + b_t^2 + \frac{e_t^2}{X_{t-1}^2} + 2(\varphi-\hat{\varphi})\, b_t + 2(\varphi-\hat{\varphi}) \frac{e_t}{X_{t-1}} \\
&\qquad + 2\frac{b_t e_t}{X_{t-1}} + \bar{b}^2 + 2\bar{b}\left((\varphi-\hat{\varphi}) + b_t + \frac{e_t}{X_{t-1}}\right)\Bigg) \mathbb{1}_{\{|X_{t-1}|\geqslant M\}} \\
&= \frac{1}{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}} \frac{1}{n}\sum_{t=1}^n b_t^2\, \mathbb{1}_{\{|X_{t-1}|\geqslant M\}} + o_P(1) \\
&= \frac{\omega^2\left((1-F(M)) + F(-M)\right) + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)}{(1-F(M)) + F(-M) + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)} + o_P(1) \\
&= \omega^2 + o_P(1).
\end{aligned}$$

This completes the proof for the second order moment, the assertion about the fourth order moment follows by a similar argumentation. □

These results immediatly can be used in the

Proof of Theorem 3.3.4. The same argumentation as in the proof of Theorem 2.1.2 gives us together with Lemmas 3.7.2 and 3.7.3:

$$\mathbf{E}^{*}\left[X_t^{*2}\right] = \frac{\mathbf{E}^{*}\left[e_{t-i}^{*2}\right]}{1-\hat{\varphi}-\mathbf{E}^{*}\left[b_t^{*2}\right]} = \frac{\sigma^2}{1-\varphi^2-\omega^2} + o_P(1)$$

A similar argumentation applies for the fourth order moment. □

As a last step of preparation to prove the validity of the bootstrap procedures, we introduce the truncated bootstrap process in a similar way to the truncated process in the standard case (Definition 2.1.5) and give a similar result.

Definition 3.7.4. *Analog to Definition 2.1.5 define the truncated version of the Bootstrap RCA process X_t^* by*

$$\tilde{X}_t^{*_s} = \sum_{i=0}^{s-1} \left(\prod_{j=0}^{i-1} (\hat{\varphi} + b_{t-j}^*) \right) e_{t-j}^*.$$

Lemma 3.7.5. *If Equation (3.2) holds, the L_2-norm of the difference between the original process X_t^* and the truncated process $\tilde{X}_t^{*_s}$ decreases to zero exponentially with increasing cut off s of $\tilde{X}_t^{*_s}$, i.e. with going further in the past:*

$$\boldsymbol{E}^* \left[\left(X_t^* - \tilde{X}_t^{*_s} \right)^2 \right] = \frac{\hat{\sigma}^2}{1 - \hat{\varphi}^2 - \hat{\omega}^2} \left(\hat{\varphi}^2 + \hat{\omega}^2 \right)^s = C \hat{\vartheta}_n^s, \ \hat{\vartheta}_n < 1 \quad i.P.$$

The constants $\left\{ \hat{\vartheta}_n^s, \ s \in \mathbb{N} \right\}$ are absolutely summable in probability.

Proof. Similar computations as in the proof of Lemma 2.1.6. $\square$

Let us now consider the asymptotic behavior of the bootstrap estimator. We can split up the difference between the bootstrap estimator and the estimator for φ as follows:

$$\hat{\varphi}^* - \hat{\varphi} = \frac{\sum X_{t-1}^* X_t^*}{\sum X_{t-1}^{*^2}} - \frac{\sum X_{t-1}^{*^2}}{\sum X_{t-1}^{*^2}} \hat{\varphi} = \frac{\sum X_{t-1}^* \left(X_t^* - \hat{\varphi} X_{t-1}^* \right)}{\sum X_{t-1}^{*^2}} = \frac{\frac{1}{n} \sum X_{t-1}^{*^2} b_t^* + X_{t-1}^* e_t^*}{\frac{1}{n} \sum X_{t-1}^{*^2}}. \tag{3.16}$$

Clearly,

$$\mathbf{Cov} \left[X_{t-1}^{*^2} b_t^*, X_{t-1}^* e_t^* \right] = \mathbf{E} \left[X_{t-1}^{*^2} b_t^* X_{t-1}^* e_t^* \right] = \mathbf{E} \left[X_{t-1}^{*^3} \right] \mathbf{E} \left[b_t^* \right] \mathbf{E} \left[e_t^* \right] = 0.$$

Instead of considering the expression of Equation (3.16) as a whole and prooving

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\frac{1}{n} \sum_{t=1}^n X_{t-1}^{*^2}} \begin{pmatrix} X_{t-1}^* e_t^* \\ X_{t-1}^{*^2} b_t^* \end{pmatrix} \xrightarrow{\mathrm{D}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{pmatrix} \right)$$

we restrict for easier notation to a proof of asymptotic normality of $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{X_t^* e_t^*}{\frac{1}{n} \sum_{t=1}^n X_{t-1}^{*^2}}$. The asymptotic normality of the second component and of the joint distribution follows by a similar argumentation.

Lemma 3.7.6. *For the residual based bootstrap the following central limit theorem holds true if Equation (3.3) is valid:*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{t-1}^* e_t^* \sqrt{\boldsymbol{E}^* \left[X_0^{*^2} \right]} \left(\frac{1}{n} \sum_{t=1}^n X_{t-1}^{*^2} \sqrt{\boldsymbol{E}^* \left[e_1^{*^2} \right]} \right)^{-1} \xrightarrow{D} \mathcal{N}(0,1)$$

Proof. For computational ease, we just consider the asymptotic behavior of

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1}^{*} e_t^{*}. \tag{3.17}$$

The assertion then easily follows with the Lemma of Slutsky and the fact that

$$\frac{1}{n} \sum_{t=1}^{n} X_{t-1}^{*^2} = \mathbf{E}^{*}\left[X_0^{*^2}\right] + o_{P^*}(1)$$

The central limit theorem for weak dependent random variables (Neumann & Paparoditis (2008), Theorem 6.1) yields the desired convergence (Eq. (3.17)) if we can show that the prerequisites are met.

Variance: $\mathbf{E}^{*}\left[X_0^{*^2} e_1^{*^2}\right] = \mathbf{E}\left[X_0^{*^2}\right] \mathbf{E}^{*}\left[e_1^{*^2}\right]$

Autocovariances: $\mathbf{E}^{*}\left[X_0^{*} e_1^{*} X_{t-1}^{*} e_t^{*}\right] = \mathbf{E}^{*}\left[X_0^{*} e_1^{*} X_{t-1}^{*}\right] \mathbf{E}\left[e_t^{*}\right] = 0$

Lindeberg condition: Let $\delta > 0$.

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \mathbf{E}^{*}\left[X_{t-1}^{*^2} e_t^{*^2} \mathbb{1}_{\{|X_{t-1}^{*} e_t^{*}| > \delta\sqrt{n}\}}\right] &\leqslant \frac{1}{n} \sum_{t=1}^{n} \mathbf{E}^{*}\left[X_{t-1}^{*^4} e_t^{*^4}\right]^{\frac{1}{2}} \mathbf{E}^{*}\left[\mathbb{1}_{\{|X_{t-1}^{*} e_t^{*}| > \delta\sqrt{n}\}}\right]^{\frac{1}{2}} \\
&\leqslant \frac{1}{n} \sum_{t=1}^{n} \mathbf{E}^{*}\left[X_{t-1}^{*^4}\right]^{\frac{1}{2}} \mathbf{E}^{*}\left[e_t^{*^4}\right]^{\frac{1}{2}} \frac{1}{\delta^2 n} \mathbf{E}^{*}\left[X_{t-1}^{*^2}\right]^{\frac{1}{2}} \mathbf{E}^{*}\left[e_t^{*^2}\right]^{\frac{1}{2}} \\
&= \frac{1}{\delta^2 n} \left(\mathbf{E}^{*}\left[X_0^{*^4}\right] \mathbf{E}^{*}\left[e_1^{*^4}\right] \mathbf{E}^{*}\left[X_0^{*^2}\right] \mathbf{E}^{*}\left[e_1^{*^2}\right]\right)^{\frac{1}{2}} \xrightarrow{\text{i.P.}} 0
\end{aligned}$$

Weak-dependance conditions: Let f be measurable and square-integrable and $s_1 < \cdots < s_u < s_u + r = t_1 \leqslant t_2$. With Lemma 3.7.5 we obtain:

$$\begin{aligned}
&\mathbf{Cov}^{*}\left[f(X_{s_1-1}^{*} e_{s_1}^{*}, \ldots, X_{s_u-1}^{*} e_{s_u}), X_{t_1-1}^{*} e_{t_1}^{*}\right] \\
&\quad = \mathbf{Cov}^{*}\left[f(X_{s_1-1}^{*} e_{s_1}^{*}, \ldots, X_{s_u-1}^{*} e_{s_u}), X_{t_1-1}^{*} e_{t_1}^{*} - \tilde{X}_{t_1-1}^{*t_1-s_u} e_{t_1}^{*}\right] \\
&\quad \leqslant \mathbf{E}^{*}\left[f^2(X_{s_1-1}^{*} e_{s_1}^{*}, \ldots, X_{s_u-1}^{*} e_{s_u})\right]^{\frac{1}{2}} \mathbf{E}^{*}\left[\left(X_{t_1-1}^{*} - \tilde{X}_{t_1-1}^{*t_1-s_u}\right)^2 e_{t_1}^{*^2}\right]^{\frac{1}{2}} \\
&\quad = \mathbf{E}^{*}\left[f^2(X_{s_1-1}^{*} e_{s_1}^{*}, \ldots, X_{s_u-1}^{*} e_{s_u})\right]^{\frac{1}{2}} \left(\frac{\hat{\sigma}_n^2}{1 - \hat{\varphi}_n^2 - \hat{\omega}_n^2} \left(\hat{\varphi}_n^2 + \hat{\omega}_n^2\right)^r\right)^{\frac{1}{2}} \\
&\quad = \kappa\, \vartheta_{n,1}^{r}.
\end{aligned}$$

with the argumentation of Remark 3.3.3 it can easily be seen that the series $\{\hat{\vartheta}_{n,1}^{r},\ r \in \mathbb{N}\}$ is absolutely summable in probability: Since $\hat{\vartheta}_{n,1} \xrightarrow{n\to\infty} \vartheta < 1$ *i.P.*, it is clear that for an arbitrary high probability and for $n > n_0$ large enough $|\hat{\vartheta}_{n,1}| \leqslant \tilde{\vartheta} < 1$. For all $n \leqslant n_0$ we rescale the estimators such that $\hat{\vartheta}_n < 1$. Therefore, $\sum_{r=1}^{\infty} \hat{\vartheta}_{n,1}^{r} \leqslant \frac{1}{1-\tilde{\vartheta}}$ *i.P.* uniformly in n.

This is not exactly the condition needed for the theorem of Neumann & Paparoditis (2008). However, the theorem does also work for the sequences $\hat{\vartheta}^r_{n,1}$ that we have here in probability. Taking a look at the proof of the theorem, it can be seen that the absolute summability of the sequences $\hat{\vartheta}^r_{n,1}$ is essential. Furthermore, the fact that $\hat{\vartheta}^r_{n,1}$ is bounded is needed, but this is fulfilled automatically with the first condition being met.
Let now f be measurable and bounded and $s_1 < \cdots < s_u < s_u + r = t_1 \leqslant t_2$. With Lemma 3.7.5 we obtain:

$$
\begin{aligned}
&\mathbf{Cov}^*\left[f(X^*_{s_1-1}e^*_{s_1},\dots,X^*_{s_u-1}e_{s_u}),X^*_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2}\right]\\
&\quad= \mathbf{Cov}^*\left[f(X^*_{s_1-1}e^*_{s_1},\dots,X^*_{s_u-1}e_{s_u}),X^*_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2} - \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}\tilde{X}^{*t_2-s_u}_{t_2-1}e^*_{t_2}\right]\\
&\quad\leqslant \|f\|_\infty \mathbf{E}^*\left[|X^*_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2} - \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}\tilde{X}^{*t_2-s_u}_{t_2-1}e^*_{t_2}|\right]\\
&\quad= \|f\|_\infty \mathbf{E}^*\Big[|X^*_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2} - \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2}\\
&\qquad\qquad\qquad\qquad + \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}X^*_{t_2-1}e^*_{t_2} - \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}\tilde{X}^{*t_2-s_u}_{t_2-1}e^*_{t_2}|\Big]\\
&\quad\leqslant \|f\|_\infty\Bigg(\left(\mathbf{E}^*\left[\left(X^*_{t_1-1}e^*_{t_1} - \tilde{X}^{*t_1-s_u}_{t_1-1}e^*_{t_1}\right)^2\right]\mathbf{E}^*\left[X^{*^2}_{t_2-1}e^{*^2}_{t_2}\right]\right)^{\frac{1}{2}}\\
&\qquad\qquad + \left(\mathbf{E}^*\left[\left(X^*_{t_2-1}e^*_{t_2} - \tilde{X}^{*t_2-s_u}_{t_2-1}e^*_{t_2}\right)^2\right]\mathbf{E}^*\left[\tilde{X}^{*^2_{t_1-s_u}}_{t_1-1}e^{*^2}_{t_1}\right]\right)^{\frac{1}{2}}\Bigg)\\
&\quad\leqslant \|f\|_\infty\left(\frac{\hat{\sigma}^2}{1-\hat{\varphi}^2-\hat{\omega}^2}\left(\left(\hat{\varphi}^2+\hat{\omega}^2\right)^r\frac{\hat{\sigma}^2}{1-\hat{\varphi}^2-\hat{\omega}^2} + \left(\hat{\varphi}^2+\hat{\omega}^2\right)^{r+t_2-t_1}\sigma^2\mathbf{E}^*\left[\tilde{X}^{*^2_{t_2-s_u}}_{t_1-1}\right]\right)\right)^{\frac{1}{2}}\\
&\quad= \|f\|_\infty\left(\hat{\varphi}^2+\hat{\omega}^2\right)^r K\\
&\quad= \|f\|_\infty \vartheta^r_{n,2}K
\end{aligned}
$$

Analog to the previous case, the series $\{\vartheta^r_{n,2},\ r \in \mathbb{N}\}$ is absolutely summable in probability as well. The argumentation used there yields that the theorem of Neumann & Paparoditis (2008) can be applied to obtain the desired convergence in probability. $\square$

Now we are ready to complete the validity of the residual based bootstrap with the

Proof of Theorem 3.3.5. Since we have seen in Theorem 3.3.4 and Lemma 3.7.2 that

$$
\begin{aligned}
\mathbf{E}^*\left[X^{*^2}_0\right] &\overset{n\to\infty}{\longrightarrow} \mathbf{E}\left[X^2_0\right]\ i.P.\\
\mathbf{E}^*\left[X^{*^2}_0\right]\mathbf{E}^*\left[e^{*^2}_1\right] &\overset{n\to\infty}{\longrightarrow} \mathbf{E}\left[X^2_0\right]\mathbf{E}\left[e^2_1\right]\ i.P.\\
\mathbf{E}^*\left[X^{*^4}_0\right]\mathbf{E}^*\left[b^{*^2}_1\right] &\overset{n\to\infty}{\longrightarrow} \mathbf{E}\left[X^4_0\right]\mathbf{E}\left[b^2_1\right]\ i.P.
\end{aligned}
$$

the assertion follows from Lemmas 3.7.6 and the argumentation before it.

$\square$

4 Simultaneous bootstrap for all three parameters of random coefficient autoregressive models

In the previous chapter, we introduced three different bootstrap approaches that work for the distribution of the autoregressive parameter φ. However, strong moment assumptions on X_t resulting in a very limited parameter space for φ and the variance ω^2 of the disturbance noise have to be put on the process to prove validity of the bootstrap. Following Aue et al. (2006) we now introduce two bootstrap approaches that put only very weak assumptions on the noise sequences and do not need any moment assumptions on the process X_t itself. Another advantage of these approaches is that they do not only work for distribution of the parameter φ but also for the distribution of the two variances ω^2 and σ^2. The bootstrap approaches are based on the quasi-maximum likelihood estimator that was introduced in Section 2.2.1. Again, we first introduce a generalization of the classical residual bootstrap for standard autoregressive processes and present a wild bootstrap version thereafter.

4.1 Residual based simultaneous bootstrap

For the QML estimator we can use the same bootstrap idea that we introduced for the least squares estimator in the previous chapter. We recall that we used the "small" observations $|X_{t-1}| \leqslant \varepsilon$ to determine estimated residuals for the innovation noise and the "large" obersvations $|X_{t-1}| \geqslant M$ to determine estimated residuals for the disturbance noise. Hence, the bootstrap procedure now is as follows.

Bootstrap Procedure 4.1.1. *Given $X_1, \dots, X_n$ we define*

$$\Gamma = \left\{ u = (s, x, y) : \; -s_0 \leqslant s \leqslant s_0, \; \frac{1}{x_0} \leqslant x \leqslant x_0, \; \frac{1}{y_0} \leqslant y \leqslant y_0 \right\}.$$

with $s_0 > 0$, $x_0 > 1$, $y_0 > 1$ such that $\theta = (\varphi, \omega^2, \sigma^2) \in int\, \Gamma$ and estimate θ via the QML-estimator $\hat{\theta} = (\hat{\varphi}, \hat{\omega}^2 \hat{\sigma}^2)$ that is given by

$$l_n\left((\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2)\right) = \inf_{u \in \Gamma} l_n(u) = \inf_{u \in \Gamma} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \ln(xX_{i-1}^2 + y) \right).$$

Denote by $2 \leqslant T_1 < T_2 < \cdots < T_{N_e}$, $N_e = \sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}| < \varepsilon\}}$, the ordered random set of indices t for which $|X_{T_i - 1}| < \varepsilon(n)$ is fulfilled and compute the related centered residuals

$$\hat{e}_{T_i} = X_{T_i} - \hat{\varphi} X_{T_i - 1} - \frac{1}{N_e} \sum_{i=1}^{N_e} \left(X_{T_i} - \hat{\varphi} X_{T_i - 1} \right).$$

Denote by $2 \leqslant S_1 < S_2 < \cdots < S_{N_b}$, $N_b = \sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}| \geqslant M\}}$, *the ordered random set of indices* t *for which* $|X_{S_i - 1}| \geqslant M(n)$ *is fulfilled and compute the related centered residuals*

$$\hat{b}_{S_i} = \frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi} - \frac{1}{N_b} \sum_{i=1}^{N_b} \left(\frac{X_{S_i}}{X_{S_i-1}} - \hat{\varphi} \right).$$

This yields approximative residuals $\hat{e}_{T_1}, \ldots, \hat{e}_{T_{N_e}}$ *and* $\hat{b}_{S_1}, \ldots, \hat{b}_{S_{N_b}}$ *with* $N_e, N_b \leqslant n$. *Generate independent bootstrap innovations* $(e_t^*,\ t \in \mathbb{Z})$ *by drawing with replacement from the set* $\left\{ \hat{e}_{T_1}, \ldots, \hat{e}_{T_{N_e}} \right\}$ *and independent bootstrap disturbances* $(b_t^*,\ t \in \mathbb{Z})$ *by drawing with replacement from the set* $\left\{ \hat{b}_{S_1}, \ldots, \hat{b}_{S_{N_b}} \right\}$, *both independent of each other. This is the classical bootstrap idea.*
Construct bootstrap observations of the process by

$$X_t^* = (\hat{\varphi} + b_t^*)\, X_{t-1}^* + e_t^*, \quad t \in \mathbb{Z}. \tag{4.1}$$

□

Asymptotic properties which $(M(n) : n \in \mathbb{N})$ and $(\varepsilon(n) : n \in \mathbb{N})$ have to fulfill can be found in the next section. Some practical remarks on how to choose the bandwidths can be found in Section 4.5. We define the estimator for the bootstrap data similar to the QML estimator given above:

Definition 4.1.2 (Bootstrap Estimator)**.** *Given* $X_1, \ldots, X_n$ *we define*

$$\Gamma^* = \left\{ u^* = (s^*, x^*, y^*) : -s_0^* \leqslant s^* \leqslant s_0^*,\ \frac{1}{x_0^*} \leqslant x^* \leqslant x_0^*,\ \frac{1}{y_0^*} \leqslant y^* \leqslant y_0^* \right\}$$

with $s_0^* > 0$, $x_0^* > 1$, $y_0^* > 1$ *such that* $\hat{\theta} = (\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2) \in \mathit{int}\, \Gamma^*$ *and* $\Gamma^* \subset \Gamma$. *For* $u^* = (s^*, x^*, y^*)$ *define the log-likelihoodfunction (suppose* b_t, e_t *normal)*

$$l_n(X^*, u^*) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(X_i^* - s^* X_{i-1}^*)^2}{x^* X_{i-1}^{*^2} + y^*} + \ln(x^* X_{i-1}^{*^2} + y^*) \right)$$

and determine the QML-estimator $\hat{\theta}^* = \left(\hat{\varphi}^*, \hat{\omega}^{*^2}, \hat{\sigma}^{*^2} \right)$ *via*

$$l_n\left(X^*, \hat{\theta}^* \right) = \inf_{u^* \in \Gamma^*} l_n(X^*, u^*)$$

Since we know the value of $\hat{\theta}$ it is easy to choose Γ^* appropriately, for example $\Gamma^* = \Gamma$.

4.2 Consistency of the residual based bootstrap

We recall that in the previous chapter (Definition 3.3.1) we introduced a notation for convergence in probability that we use in the following as well. We first state the following assumptions that are assumed to hold throughout this section.

$$X_t \text{ has a density that is positive on the whole real axis and continuous in zero.} \tag{4.2}$$

$$(e_t)_{t \in \mathbb{Z}} \text{ and } (b_t)_{t \in \mathbb{Z}} \text{ have finite sixths moments} \tag{4.3}$$

$$\varepsilon = \varepsilon(n) \overset{n\to\infty}{\longrightarrow} 0 \text{ is chosen such that } n^2 \varepsilon^3 \overset{n\to\infty}{\longrightarrow} \infty. \tag{4.4}$$

$$M = M(n) \overset{n\to\infty}{\longrightarrow} \infty \text{ is such that } \sqrt{n}(1 - F(M)) \overset{n\to\infty}{\longrightarrow} \infty \text{ and } \sqrt{n} F(-M) \overset{n\to\infty}{\longrightarrow} \infty. \tag{4.5}$$

Lemma 4.2.1. *For all s for that the moments of e_t and b_t, respectively, exist, it holds $\boldsymbol{E}^*\left[b_t^{*^s}\right] \overset{n\to\infty}{\longrightarrow} \boldsymbol{E}[b_t^s]$ i.P., $\boldsymbol{E}^*\left[e_t^{*^s}\right] \overset{n\to\infty}{\longrightarrow} \boldsymbol{E}[e_t^s]$ i.P., and further $\boldsymbol{E}^*\left[\ln^+|e_0^*|\right] = \mathcal{O}_P(1)$, $\boldsymbol{E}^*\left[\ln^+|\hat{\varphi} + b_0^*|\right] = \mathcal{O}_P(1)$, $-\infty < \boldsymbol{E}^*\left[\ln|\hat{\varphi} + b_0^*|\right] < 0$ i.P., hence X_t^* constructed by Equation (4.1) in Procedure 4.1.1 converges absolutely in the form*

$$X_t^* = \sum_{i=0}^{\infty}\left(\prod_{j=0}^{i-1}(\hat{\varphi} + b_{t-j}^*)\right) e_{t-i}^* \tag{4.6}$$

and is strictly stationary in probability.

The proof is deferred to Section 4.6. □

Lemma 4.2.2. *For the process X_t^* holds that*

$$\boldsymbol{E}^*\left[\frac{X_{i-1}^{*^\kappa}}{\left(\hat{\omega}^2 X_{i-1}^{*^2} + \hat{\sigma}^2\right)^\gamma}\right] \xrightarrow{i.P.} \boldsymbol{E}\left[\frac{X_{i-1}^\kappa}{\left(\omega^2 X_{i-1}^2 + \sigma^2\right)^\gamma}\right], \quad 0 < \kappa < 2\gamma, \quad \gamma \in \mathbb{N}.$$

The proof is delayed to Section 4.6. □

Now we can consider the derivatives of g :

Lemma 4.2.3. *$A^* = \boldsymbol{E}^*\left[g'(X_0^*, \hat{\theta})^\top g'(X_0^*, \hat{\theta})\right]$ exists and is non-singular in probability and*

$$\boldsymbol{E}^*\left[\frac{\partial}{\partial u_j} g(X_0^*, u^*)\right] \xrightarrow{i.P.} \boldsymbol{E}\left[\frac{\partial}{\partial u_j} g(X_0, u)\right], \; j = 1, 2, 3,$$
$$A^* = \boldsymbol{E}^*\left[g'(X_0^*, \hat{\theta})^\top g'(X_0^*, \hat{\theta})\right] \xrightarrow{i.P.} \boldsymbol{E}\left[g'(X_0, \theta)^\top g'(X_0, \theta)\right] = A.$$

Further, $H^* = \boldsymbol{E}^*\left[g_1''(X_0^*, \hat{\theta})\right]$ exists and is non-singular in probability and

$$\frac{1}{n}\sum_{i=1}^{n} g''(X_i^*, \hat{\theta}) = H^* + o_{P^*}(1)$$
$$H^* = \boldsymbol{E}^*\left[g''(X_0^*, \hat{\theta})\right] \xrightarrow{i.P.} \boldsymbol{E}\left[g''(X_0, \theta)\right] = H.$$

The proof is postponed to Section 4.6. □
Our analysis of the bootstrap estimator follows the argumentation of Aue et al. (2006).

Lemma 4.2.4. *The following convergence holds true for the QML estimator $\hat{\theta}$, the data generated by the bootstrap procedure 4.1.1 and the bootstrap estimator $\hat{\theta}^*$ of Definition 4.1.2:*

$$\hat{\theta}_n^* - \hat{\theta}_n \xrightarrow{i.P.} 0 \quad i.P.$$

The proof is deferred to Section 4.6. □
With the next result we finally arrive at the main theorem of this section.

Lemma 4.2.5. *Under conditions (4.2) through (4.5) the following convergence holds true for the QML estimator $\hat{\theta}$, the data generated by the bootstrap procedure 4.1.1 and the bootstrap estimator $\hat{\theta}^*$ of Definition 4.1.2:*

$$\sqrt{n}\, l_n'(X^*, \hat{\theta}) \xrightarrow{D} \mathcal{N}(0, A) \quad i.P.$$

The proof is delayed to Section 4.6. □

Theorem 4.2.6. *Under the assumptions of Lemma 4.2.5 the following convergence holds true:*

$$\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right) \xrightarrow{D} \mathcal{N}\left(0, H^{-1}AH^{-1}\right) \quad i.P.$$

Proof. We know from Lemma 4.2.3 that H^{*-1} exists and that it is non-singular in probability. $l_n(X^*, u^*)$ is continuously differentiable on $int\,\Gamma^*$ and $\hat{\theta}^* = \hat{\theta} + o_{P^*}(1)$ *i.P.*, therefore, the likelihood function attains its maximum on $int\,\Gamma^*$ for n large enough and it follows that for n large enough $l_n'(X^*, \hat{\theta}^*) = 0$, since the estimator is constructed this way. Hence,

$$\begin{aligned}
-l_n'(X^*, \hat{\theta}) &= l_n'(X^*, \hat{\theta}^*) - l_n'(X^*, \hat{\theta}) \\
&= l_n''(X^*, \vartheta)\, \left(\hat{\theta}^* - \hat{\theta}\right) \quad \text{with } \vartheta \text{ between } \hat{\theta} \text{ and } \hat{\theta}^* \\
&= \frac{1}{n}\sum_{i=1}^{n} g''(X_i^*, \vartheta)\, \left(\hat{\theta}^* - \hat{\theta}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left(g''(X_i^*, \hat{\theta}) + (\vartheta - \hat{\theta}) g'''(X_i^*, a)\right)\, \left(\hat{\theta}^* - \hat{\theta}\right) \quad \text{with } a \text{ between } \vartheta \text{ and } \hat{\theta} \\
&= \left(H^* + o_P(1)\right) \left(\hat{\theta}^* - \hat{\theta}\right)
\end{aligned}$$

where we note that $\left|(\vartheta - \hat{\theta})\, g'''(X_i^*, a)\right| = \left|(\vartheta - \hat{\theta})\right| \left|g'''(X_i^*, a)\right| \leqslant \left|\hat{\theta}^* - \hat{\theta}\right| \mathcal{O}_P(1) = o_P(1)$ if $g'''(X_i^*, a) = \mathcal{O}_P(1)$, what follows from the proof of Lemma 8 in Aue et al. (2006) and Lemma 4.2.3 above. Hence, the assertion follows with Slutsky from Lemmas 4.2.3 and 4.2.5. □

Corollary 4.2.7. *Under the assumptions of Lemma 4.2.5 the bootstrap of Procedure 4.1.1 is consistent for $\theta = (\varphi, \omega^2, \sigma^2)$.*

Proof. If we denote the distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ by $\mathcal{L}_n$ and the one of $\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$ by $\mathcal{L}_n^*$, it directly follows that

$$d_K\left(\mathcal{L}_n, \mathcal{L}_n^*\right) = \sup_{x \in \mathbb{R}} \left|\mathcal{L}_n\{(-\infty, x]\} - \mathcal{L}_n^*\{(-\infty, x]\}\right| \xrightarrow{n \to \infty} 0 \quad i.P.$$

by applying Theorems 2.2.2 and 4.2.6 and the triangle inequality.

□

4.3 Density based simultaneous bootstrap

For the remainder of the chapter, we assume that the white noises $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t\in\mathbb{Z}}$ have densities $h(\cdot)$ or $k(\cdot)$, respectively. Taking a look at the asymptotic distribution of the QML estimator in Theorem 2.2.2, it can be seen that a bootstrap method needs to capture a quite strong dependency structure of the process to mimic the variance of the asymptotic distribution correctly. The bootstrap approach we just presented is, however, based on possibly very few "small" and "large" observations, so a limited number of random numbers is repeated frequently to obtain bootstrap realizations. Hence, one might think that one can obtain better results using a smoothed density estimator based on these values and thus enlarging the number of possible values for the bootstrap realizations of the noises. This can be done in two ways: One is to follow the idea described in Section 4.1 in using the "small" and the "large" observations of X_t to obtain estimated residuals for the innovation and the disturbance noise, respectively, and then to smooth these directly, wheras the other one is to first estimate the innovation noise and then to use deconvolution methods to obtain an estimator for the disturbance noise. This means for the first idea:

Definition 4.3.1. *In an RCA model, with $\hat{u}_t^n = X_t - \hat{\varphi}X_{t-1}$, estimators for the densities of the innovation and the disturbance noises are given by*

$$\hat{k}_{n,\varepsilon}(y) = \frac{1}{\sum_{t=1}^{n} \frac{1}{\varepsilon} W\left(\frac{X_{t-1}}{\varepsilon}\right)} \sum_{t=1}^{n} \frac{1}{h\varepsilon} K\left(\frac{\hat{u}_t^n - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)$$

$$\hat{h}_{n,M}(y) = \frac{1}{\sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}} \sum_{t=1}^{n} \frac{1}{h_1} G\left(\frac{\frac{X_t}{X_{t-1}} - \hat{\varphi} - y}{h_1}\right) \mathbb{1}_{\{|X_{t-1}|\geqslant M\}}$$

where K, G, and W are kernel functions and the bandwidths are $h = h(n) \xrightarrow{n\to\infty} 0$, $h_1 = h_1(n) \xrightarrow{n\to\infty} 0$, and $\varepsilon = \varepsilon(n) \xrightarrow{n\to\infty} 0$ and a parameter $M = M(n) \xrightarrow{n\to\infty} \infty$.

This approach again has the drawback that it is not clear, which observations are large enough to obtain reliable results and if there are enough "large" observations. This should be of less importance when using the deconvolution methods because for this, one could use, for example, all observations of the process X_t near one to estimate the sum $u_t = b_t + e_t$ and extract the density of b_t from this. We note that for the characteristic functions we have the coherence $\phi_u = \phi_b \cdot \phi_e$, where we denote by ϕ_Y the characteristic function of a random variable Y. Replacing the characteristic functions by estimates we obtain:

Definition 4.3.2. *An estimator for the density of the disturbance parameter is given by*

$$\begin{aligned}
\hat{h}_n(z) &= \frac{1}{2\pi} \int \mathrm{e}^{-itz} \frac{\hat{\phi}_u(t)}{\hat{\phi}_e(t)} dt \\
&= \frac{1}{n\delta} \sum_{j=1}^{n} \frac{1}{2\pi} \int e^{it(\hat{u}_j^n - z)} \frac{\phi_G(tk) V\left(\frac{X_{j-1}-a}{\delta}\right)}{\frac{1}{n\varepsilon}\sum_{r=1}^{n} \mathrm{e}^{it\hat{u}_r^n} \phi_K(th) W\left(\frac{X_{t-1}}{\varepsilon}\right)} dt \, \frac{\frac{1}{n\varepsilon}\sum_{r=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)}{\frac{1}{n\delta}\sum_{r=1}^{n} V\left(\frac{X_{t-1}-1}{\delta}\right)},
\end{aligned}$$

where $\hat{\varphi}$ is an estimator for φ, $\hat{u}_t^n = X_t - \hat{\varphi}X_{t-1}$, K, G, V, W are kernel functions and h, k, ε, $\delta \xrightarrow{n\to\infty} 0$ are bandwidths in dependence on n. This estimator can also be generalized to using any point a intead of 1 to estimate the sum $ab_t + e_t$ to extract the density of b_t from. □

These three density estimators and some variations of them will be considered in more detail in the next chapter. For the bootstrap procedure, one could of course think of other ways to estimate these densities since we only need estimators that exhibit the following characteristics and that include the estimators just mentioned.

Definition 4.3.3. *Assume that we are given a consistent estimator $\hat{\varphi}$ for φ as well as estimators $\hat{k}_n(x)^+$ and $\hat{h}_n(x)^+$ for the densities of e_t and b_t with*

$$\sup_{x\in[-R,R]} \left|\hat{k}_n(x)^+ - k(x)\right| = \mathcal{O}_P(a_n) \qquad \textit{and} \qquad \sup_{x\in[-R,R]} \left|\hat{h}_n(x)^+ - h(x)\right| = \mathcal{O}_P(b_n)$$

for sequences $(a_n)_{n\in\mathbb{N}} \overset{n\to\infty}{\longrightarrow} 0$ and $(b_n)_{n\in\mathbb{N}} \overset{n\to\infty}{\longrightarrow} 0$. For the following, we define density estimators that are truncated to zero outside compact intervals $[-R, R]$ slowly growing to the real axis by

$$\hat{k}_n(x) = \hat{k}_n(x)^+ \, \mathbb{1}_{[-R,R]}(x) \qquad \textit{and} \qquad \hat{h}_n(x) = \hat{h}_n(x)^+ \, \mathbb{1}_{[-R,R]}(x).$$

Clearly, these estimators are also uniformly consistent on $[-R, R]$ for the densities for e_t and b_t, respectively. For practical implementation, these estimators might have to be normed since they do not necessarily integrate to one. Furthermore, we have

Lemma 4.3.4. *Let $\hat{k}(\cdot)$ and $\hat{h}(\cdot)$ be estimators for $k(\cdot)$ and $h(\cdot)$ as given in Definition 4.3.3. Then the moments of order s of e_t and b_t can be estimated consistently by*

$$\hat{\boldsymbol{E}} e_t^s := \int x^s \hat{k}_n(x)\, dx \overset{i.P.}{\longrightarrow} \boldsymbol{E}\left[e_t^s\right] \qquad \textit{and} \qquad \hat{\boldsymbol{E}} e_t^s := \int x^s \hat{h}_n(x)\, dx \overset{i.P.}{\longrightarrow} \boldsymbol{E}\left[b_t^s\right]$$

for all $s \in \mathbb{N}$ for that the moments of e_t or b_t, respectively, exist, if $R = R(n) \overset{n\to\infty}{\longrightarrow} \infty$ is chosen such that $R^s a_n \overset{n\to\infty}{\longrightarrow} 0$, and also for all $a \leqslant b$:

$$\int_a^b \hat{k}_n(x)\, dx \overset{i.P.}{\longrightarrow} \int_a^b k_n(x)\, dx \qquad \textit{and} \qquad \int_a^b \hat{h}_n(x)\, dx \overset{i.P.}{\longrightarrow} \int_a^b h_n(x)\, dx.$$

Proof. The assertions follow by direct computations. Consider exemplary the first one:

$$\begin{aligned}
\int x^s \hat{k}_n(x)\mathrm{dx} &= \int_{-R}^{R} x^s \left(k(x) + \mathcal{O}_P\left(a_n\right)\right) \mathrm{dx} \\
&= \int_{-R}^{R} x^s k(x)\mathrm{dx} + \mathcal{O}_P(a_n) \int_{-R}^{R} x^s C \mathrm{dx} \quad \text{with } C \text{ independent of } x \\
&= \mathbf{E}[e_t^s] + o(1) + \mathcal{O}_P(a_n R^s)
\end{aligned}$$

□

Remark 4.3.5. *Typically, a_n and b_n are of polynomial order, hence a possible choice for R would be a logarithmic order.*

With that in mind we set up a variation of the residual based bootstrap:

Bootstrap Procedure 4.3.6. *With $s_0 > 0,\ x_0 > 1,\ y_0 > 1$ such that $\theta = (\varphi, \omega^2, \sigma^2) \in \operatorname{int}\Gamma$ we define $\Gamma = \left\{u = (s, x, y) :\ -s_0 \leqslant s \leqslant s_0,\ \frac{1}{x_0} \leqslant x \leqslant x_0,\ \frac{1}{y_0} \leqslant y \leqslant y_0\right\}$ and estimate θ via the QML-estimator $\hat{\theta} = (\hat{\varphi}, \hat{\omega}^2 \hat{\sigma}^2)$ that is given by*

$$l_n\left((\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2)\right) = \inf_{u\in\Gamma} l_n(u) = \inf_{u\in\Gamma} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \ln(xX_{i-1}^2 + y) \right). \tag{4.7}$$

Assume that we are given consistent estimators $\hat{h}_n(x)$ and $\hat{k}_n(x)$ for the densities of b_t and e_t, with the characteristics as given in Definition 4.3.3 and Lemma 4.3.4. Additionally, we require that the densities are normed to mean zero and variance $\hat{\omega}^2$ and $\hat{\sigma}^2$, respectively. Generate independant bootstrap observations $b_t^* \sim \hat{h}(x),\ t \in \mathbb{Z}$, and $e_t^* \sim \hat{k}(x),\ t \in \mathbb{Z}$. Construct bootstrap observations of the RCA process by

$$X_t^* = X_{t,n,r}^* = \sum_{i=0}^{r-1} \left(\prod_{j=0}^{i-1} (\hat{\varphi} + b_{t-j}^*) \right) e_{t-i}^*, \quad t \in \mathbb{Z}. \tag{4.8}$$

with $r = r(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ such that $\hat{f}_{n,r}(\cdot)$ as given in Definition 4.4.1 is a consistent estimator for $f(\cdot)$. □

As a bootstrap estimator, we use the estimator given in Definition 4.1.2.

4.4 Consistency of the density based bootstrap

To show that the bootstrap procedure 4.3.6 is consistent for all three parameters $(\varphi, \omega^2, \sigma^2)$ we first consider the problem of estimating the densities consistently in more detail and recall that in Definition 2.1.5 we defined the truncated version of an RCA process X_t by

$$\tilde{X}_t^r = \sum_{i=0}^{r-1} \left(\prod_{j=0}^{i-1} (\varphi + b_{t-j}) \right) e_{t-i}.$$

Lemma 4.4.1. *With $\hat{h}(\cdot)$ and $\hat{k}(\cdot)$ as given in Definition 4.3.3, for $r > 0$ and $R = R(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ such that $R^{\frac{(r+1)(r+2)}{2}-1} \max(a_n, b_n) \stackrel{n\to\infty}{\longrightarrow} 0$, an estimator for the density of the truncated process $\tilde{X}_t^r$ from Definition 2.1.5 is given by*

$$\hat{f}_{n,r}(x) := \prod_{i=1}^{r} \prod_{j=1}^{r-1} \int \cdots \int \hat{h}_n(x_i - \varphi) \hat{k}_n\left(\frac{y_r}{x_r}\right) \hat{k}_n\left(\frac{y_j}{x_j} - x_{j+1}\right) \hat{k}_n(x - y_1)\, dx_1 \ldots dx_r\, dy_1 \ldots dy_r,$$

with $\hat{f}_{n,r}(x) = f_r(x) + o_P(1)$ for all $x \in \left[-R/\left(\frac{(r+1)(r+2)}{2} - 1\right),\ R/\left(\frac{(r+1)(r+2)}{2} - 1\right)\right]$. Outside of this interval the estimator is defined to be zero.

The proof is postponed to Section 4.6. □

Theorem 4.4.2. *With $\hat{h}(\cdot)$ and $\hat{k}(\cdot)$ as given in Definition 4.3.3, the density $\hat{f}_{n,r}(\cdot)$ as given in Lemma 4.4.1 is also an estimator for the density $f(\cdot)$ of X_t if $R = R(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ and $r = r(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ are such that $R^{\frac{(r+1)(r+2)}{2}-1} \max(a_n, b_n) \stackrel{n\to\infty}{\longrightarrow} 0$ and $R/\left(\frac{(r+1)(r+2)}{2} - 1\right) \stackrel{n\to\infty}{\longrightarrow} \infty$ and we have the result*

$$\hat{f}_{n,r}(x) \stackrel{n\to\infty}{\longrightarrow} f(x) \quad i.P. \quad \forall\, x \in \mathbb{R}.$$

The proof can be found in Section 4.6. □

We next note that the moments of the bootstrap random variables converge to the moments of the real world variables as desired:

Lemma 4.4.3. *For all s for which the s-th moments of $(e_t)_{t\in\mathbb{Z}}$ or $(b_t)_{t\in\mathbb{Z}}$, respectively, exist, it holds that $\boldsymbol{E}^*\left[b_t^{*^s}\right] \xrightarrow{n\to\infty} \boldsymbol{E}[b_t^s]$ i.P. and $\boldsymbol{E}^*\left[e_t^{*^s}\right] \xrightarrow{n\to\infty} \boldsymbol{E}[e_t^s]$ i.P. and further that $\boldsymbol{E}^*\left[\ln^+|e_0^*|\right] = \mathcal{O}_P(1)$ and $\boldsymbol{E}^*\left[\ln^+|\hat{\varphi}+b_0^*|\right] = \mathcal{O}_P(1)$ and $-\infty < \boldsymbol{E}^*\left[\ln|\hat{\varphi}+b_0^*|\right] \leqslant 0$ i.P-Pr.. Hence, X_t^* as given in Equation (4.8) in Procedure 4.3.6 converges absolutely in probability and is strictly stationary in probability and it has a density $\hat{f}_{n,r}$ as given in Lemma 4.4.1 in probability. Further,*

$$\begin{aligned}
\boldsymbol{E}^*\left[X_t^*\right] &= \boldsymbol{E}\left[X_t\right], \\
\boldsymbol{E}^*\left[X_t^{*^2}\right] &\xrightarrow{i.P.} \boldsymbol{E}\left[X_t^2\right] \quad \text{if } \boldsymbol{E}\left[X_t^2\right] < \infty, \\
\boldsymbol{E}^*\left[\frac{X_{i-1}^{*^\kappa}}{\left(\hat{\omega}^2 X_{i-1}^{*^2}+\hat{\sigma}^2\right)^\gamma}\right] &\xrightarrow{i.P.} \boldsymbol{E}\left[\frac{X_{i-1}^{\kappa}}{\left(\omega^2 X_{i-1}^{2}+\sigma^2\right)^\gamma}\right] \quad \kappa = 1,\ldots,2\gamma,\ \gamma\in\mathbb{N}.
\end{aligned}$$

The proof is suspended to Section 4.6. □

Now we can conclude on the consistency of Bootstrap Procedure 4.3.6:

Theorem 4.4.4. *The following convergence holds true for the estimator $\hat{\theta}$ given in Equation (2.7), the data generated by Procedure 4.3.6, and the bootstrap estimator $\hat{\theta}^*$ given in Definition 4.1.2:*

$$\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right) \xrightarrow{D} \mathcal{N}\left(0, H^{-1}AH^{-1}\right) \ i.P.,$$

hence, the bootstrap of Procedure 4.3.6 is consistent for $\theta = (\varphi, \omega^2, \sigma^2)$.

Proof. The same statements as in Lemma 4.2.3 can be established for this bootstrap procedure so that it follows analogously to Lemma 4.2.4 that $\hat{\theta}_n^* - \hat{\theta}_n \xrightarrow{\text{i.P.}} 0$ *i.P.* and analogously to Lemma 4.2.5 that $\sqrt{n}\, l_n'(X^*, \hat{\theta}) \xrightarrow{\text{D}} \mathcal{N}(0, A)$ *i.P.*, so that the desired convergence can be obtained analogously to Theorem 4.2.6. Denoting the distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ by $\mathcal{L}_n$ and the one of $\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$ by $\mathcal{L}_n^*$ as well as applying Theorem 2.2.2 and the triangular inequality directly yields that $d_K\left(\mathcal{L}_n, \mathcal{L}_n^*\right) \xrightarrow{n\to\infty} 0$ *i.P.* and thus the consistency. □

4.5 A simulation study

We choose the parameters to be $(\varphi, \omega^2, \sigma^2) = (0.55,\ 0.6,\ 0.8)$, that means parameter set II, and again b_t normally distributed and e_t double exponentially distributed. Given a sequence of observations of X, say $X_0, \ldots, X_n$, we estimate the parameter vector $\hat{\theta} = (\hat{\varphi}, \hat{\omega}^2, \hat{\sigma}^2)$ via the QML estimator given in Equation (2.7). As a initial value for the optimization we use the value of the least squares estimator given in Section 2.2.1. Even though consistency of this estimator cannot be established for our process having infinte fourth moments, it might be a good guess to start with.

Then, we generate bootstrap realizations of the process $X_0^*, \dots, X_n^*$ via the residual based bootstrap (Procedure 4.1.1 and Definition 4.1.2) to determine the bootstrap estimates $\hat{\theta}^* = (\hat{\varphi}^*, \hat{\omega}^{*^2}, \hat{\sigma}^{*^2})$. Again, the least squares estimator is used to provide a initial value for the optimization and the parameters are chosen to be $\varepsilon = 0.3n^{-1/6}$ and $M = 1.6 \ln\ln n$. This is repeated $N = 10\,000$ times to determine the α and $(1-\alpha)$ quantiles of the bootstrap distribution $\mathcal{L}(\sqrt{n}(\hat{\theta}^* - \hat{\theta}))$ for various values of α.
Finally, we also determine the respective quantiles of $\mathcal{L}(\sqrt{n}(\hat{\theta} - \theta))$ via the normal approximation given in Theorem 2.2.2. As explained before, we are faced with the problem of estimating the moments that appear in the asymptotic variance. We have estimators for the second moments of both the noises, but not for the fourth moments. Hence, one possibility is to assume that the fourth moments are equal to $4\omega^4$ or $6\sigma^4$. However, in doing so we provide very useful information for the normal approximation that usually is not available. Another possibility is to use the estimators introduced in Theorem 2.2.3 to estimate fourth moments, even though consisteny is not established without assuming finite eight moments of X_t. The term of Equation (2.9) can be estimated simply from the data.

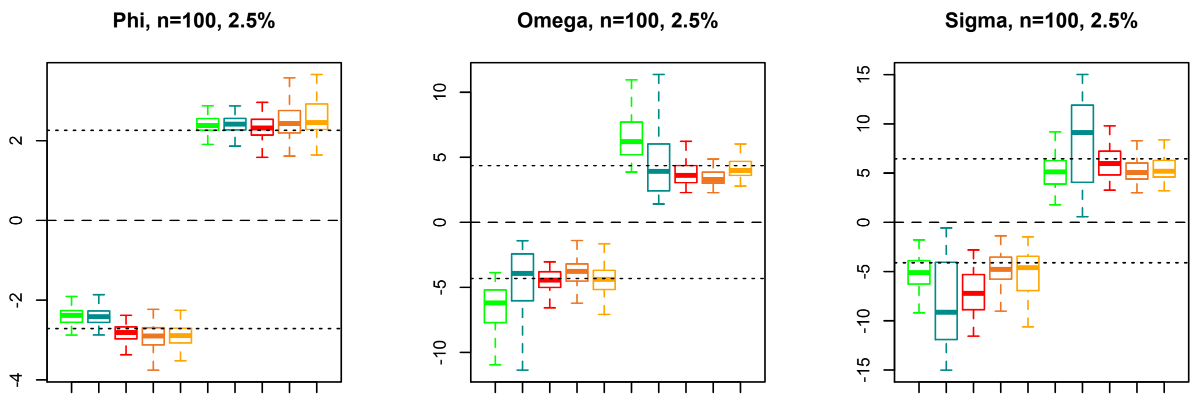

Figure 4.1: Quantiles: Bootstrap and normal approximation, repetitions $T = 100$, sample size $n = 100$

To this end, all this is repeated $T = 100$ times to obtain the boxplots that are shown in Figure 4.1. The dotted lines give the true (obtained by simulations) intervals, the blue and green boxplots give the approximations via the normal distribution (green: moment relations, blue: moment estimators), while the red boxplots give the approximations obtained by the residual based bootstrap. In addition, we also perform the density based bootstrap. First, we just smooth the estimated residuals by the estimators given in Definition 4.3.1 with parameters set to be $h = h_1 = 0.3n^{-1/6}$, $\varepsilon = 1.8n^{-1/6}$, and $M = 1.6 \ln\ln n$. Second, we use the deconvolution estimator given in Definition 4.3.2 with additional parameters set to be $k = 1.3n^{-1/10}$ and $\delta = 3.0n^{-1/10}$. These parameters turned out to be good parameters for estimating the densities, what will be considered further in Chapter 5. However, the procedure does not seem to be very sensitive on the choice of the parameters, since the results did not change substantially when altering the parameters. Practically, we determine a discretized density on a fine grid with stepsize $s = 0.01$. The results of these two bootstrap approaches are displayed in the orange and yellow boxplot.
As one can see, all bootstrap approaches work well. For all three parameters, the bootstrap

is able to capture the skewness of the distribution, in contrast to the normal approximation. While the difference is not large for the AR parameter φ, the bootstrap outperforms the normal approximation for the two variance parameters. The residual based bootstrap already works very well, even though only very few observations are used. Some improvement is possible by smoothing the observations and using the density based bootstrap, but the improvement is neither large nor given in all situations. Hence, taking into account the much higher computational effort for the density based bootstrap, the residual based bootstrap is a reasonable proposal. The normal approximations indeed offer some drawbacks: Especially the distributions of the variances are skewed and with assuming the relationship between the fourth and second moments as indicated above the upper quantile for σ^2 is heavily underestimated while the other quantiles are overestimated. Using the fourth moment estimators lets us heavily overestimate the lower quantile, coming together with a very high variability in the estimated quantiles. Hence, the bootstrap seems to be a very good alternative in this case.

4.6 Proofs

Before we show the validity of the results, we state two useful lemmas that are valid for both of the bootstrap approaches. We recall from Definition 2.1.5 the truncated RCA process. For this process we can state

Lemma 4.6.1. *$\exists\, r_0 \in \mathbb{N}$ such that $\forall\, r > r_0$ the difference between X_t and its truncated version is bounded: $\left|X_t - \tilde{X}_t^r\right| \leqslant \rho^r C$ a.s. The same holds true for the bootstrap process regardless which of the two methods described before is used: $\left|X_t^* - \tilde{X}_t^{*^r}\right| \leqslant \tilde{\rho}^r \tilde{C}$ a.s. i.P. Moreover, for r sufficiently large and $\varepsilon > 0$, $\left|F_{\tilde{X}_t^r}(x) - F_{X_t}(x)\right| \leqslant \varepsilon$ a.s.*

Proof. From the assumptions (Definition 2.1.1 and Inequalities (2.2)) and Aue et al. (2006), Lemma 1, it follows that $\exists\, i_0,\ 0 \leqslant \rho < 1$, such that $\forall\, i > i_0 : \ \prod_{j=1}^{i} |\varphi + b_{t-j}| \leqslant \rho^i$ *a.s.* and that for r sufficiently large

$$\begin{aligned}
\left|X_t - \tilde{X}_t^r\right| &\leqslant \sum_{i=r}^{\infty} \left|\prod_{j=0}^{i-1} (\varphi + b_{t-j})\right| |e_{t-i}| \leqslant \sum_{i=r}^{\infty} \rho^i \, |e_{t-i}| \quad a.s. \\
&= \rho^r \sum_{i=0}^{\infty} \rho^i \, |e_{t-r-i}| \leqslant \rho^r C \quad a.s. \quad \text{with } C < \infty \text{ (Berkes et al. (2003), Le. 2.2).}
\end{aligned}$$

The assertion for the bootstrap process follows similarly. Further, for arbitrary $\varepsilon > 0$

$$\begin{aligned}
F_{X_t}(x) = P\left\{\tilde{X}_t^r \leqslant x - (X_t - \tilde{X}_t^r)\right\} &\leqslant P\left\{\tilde{X}_t^r \leqslant x + \delta\right\} + P\left\{|X_t - \tilde{X}_t^r| > \delta\right\} \\
&\leqslant F_{\tilde{X}_t^r}(x + \delta) + \frac{\rho^r}{\delta} C' \quad a.s. \\
&\leqslant F_{\tilde{X}_t^r}(x) + \varepsilon \quad a.s.
\end{aligned}$$

for δ sufficiently small and r sufficienly large, since $\rho < 1$ and F is continuous from above. Similarly, we obtain $F_{\tilde{X}_t^r}(x) \leqslant F_{X_t}(x) + \varepsilon$ *a.s.* so that $\left|F_{\tilde{X}_t^r}(x) - F_{X_t}(x)\right| \leqslant \varepsilon$ *a.s.* □

Lemma 4.6.2. *For X_t the original process and X_t^* the bootstrap process generated with either method mentioned above and $\kappa, \gamma \in \mathbb{N}$ with $\kappa \leqslant 2\gamma$, $\exists\, r_0$ such that $\forall\, r > r_0$ and $\forall\, u$:*

$$\left|\boldsymbol{Cov}\left[\frac{X_t^\kappa}{(x^2X_t^2+y^2)^\gamma}, \frac{X_s^\kappa}{(x^2X_s^2+y^2)^\gamma}\right]\right| \leqslant C\vartheta^r$$
$$\text{and}\ \left|\boldsymbol{Cov}^*\left[\frac{X_t^{*^\kappa}}{\left(x^2X_t^{*^2}+y^2\right)^\gamma}, \frac{X_s^{*^\kappa}}{(x^2X_s^{*^2}+y^2)^\gamma}\right]\right| \leqslant \tilde{C}\tilde{\vartheta}^r\ i.P.$$

with $C, \tilde{C} < \infty$ and $\vartheta, \tilde{\vartheta} < 1$. Therefore,

$$\frac{1}{n}\sum_{i=1}^n g(X_i,u) = \boldsymbol{E}[g(X_0)] + o_P(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n g(X_i^*,u^*) = \boldsymbol{E}[g(X_0)] + o_P(1)$$

Proof. We first note that for $\kappa \in \mathbb{N}$ holds that $X_s^\kappa - \tilde{X}_s^{r^\kappa} = \left(X_s - \tilde{X}_s^r\right)\sum_{i=1}^k X_s^{k-i}\tilde{X}_s^{r^{i-1}}$. We also know that $\exists\, r_0 \in \mathbb{N}$, $\vartheta < 1$ such that $\forall\, r > r_0$: $|X_s - \tilde{X}_s^r| \leqslant \vartheta^r$ *a.s.*, so that we have for $s > t + r_0$ with $r = s - t$:

$$\begin{aligned}
&\left|\mathbf{Cov}\left[\frac{X_t^\kappa}{(x^2X_t^2+y^2)^\gamma}, \frac{X_s^\kappa}{(x^2X_s^2+y^2)^\gamma}\right]\right| \\
&= \left|\mathbf{Cov}\left[\frac{X_t^\kappa}{(x^2X_t^2+y^2)^\gamma}, \frac{X_s^\kappa}{(x^2X_s^2+y^2)^\gamma} - \frac{\tilde{X}_s^{r^\kappa}}{(x^2X_s^2+y^2)^\gamma} + \frac{\tilde{X}_s^{r^\kappa}}{(x^2X_s^2+y^2)^\gamma} - \frac{\tilde{X}_s^{r^\kappa}}{(x^2X_s^{r^2}+y^2)^\gamma}\right]\right| \\
&\leqslant \mathbf{E}\left[\frac{X_t^{2\kappa}}{(x^2X_t^2+y^2)^{2\gamma}}\right]^{\frac{1}{2}}\left(\mathbf{E}\left[\frac{(X_s^\kappa - \tilde{X}_s^{r^\kappa})^2}{(x^2X_s^2+y^2)^{2\gamma}}\right]^{\frac{1}{2}} + x^4\mathbf{E}\left[\tilde{X}_s^{r^{2\kappa}}\frac{\left(\sum_{j=0}^\gamma \binom{\gamma}{j}\left(\tilde{X}_s^{r^{2j}} - X_s^{2j}\right)\right)^2}{(x^2X_s^2+y^2)^{2\gamma}(x^2\tilde{X}_s^{r^2}+y)^{2\gamma}}\right]^{\frac{1}{2}}\right) \\
&\leqslant C\vartheta^r
\end{aligned}$$

For the second part, we recall from Equation (2.8) that

$$g(X_i,\theta) = \frac{(X_i - \varphi X_{i-1})^2}{\omega^2X_{i-1}^2+\sigma^2} + \ln(\omega^2X_{i-1}^2+\sigma^2) = \frac{(b_iX_{i-1}+e_i)^2}{\omega^2X_{i-1}^2+\sigma^2} + \ln(\omega^2X_{i-1}^2+\sigma^2).$$

By taylor expansion, $\ln\left(\omega^2X_{i-1}^2+\sigma^2\right)$ can be fractionized into terms of the form used above. Now, the assertion about $\frac{1}{n}\sum_{i=1}^n g(X_i,u)$ follows directly since

$$\begin{aligned}
&\mathbf{Var}\left[\frac{1}{n}\sum_{i=1}^n g(X_i,u)\right] \\
&\qquad = \frac{2}{n^2}\sum_{i=1}^n\sum_{j=i}^{i+r_0}\mathbf{Cov}\left[g(X_i,u),g(X_j,u)\right] + \frac{2}{n^2}\sum_{i=1}^n\sum_{j=i+r_0+1}^n\mathbf{Cov}\left[g(X_i,u),g(X_j,u)\right] \\
&\qquad \leqslant \frac{C_1r_0}{n} + \frac{C_2}{n^2}\sum_{i=1}^n\frac{\vartheta^{i+r_0+1}-\vartheta^n}{1-\vartheta} = o(1).
\end{aligned}$$

With Lemmas 4.2.1, 4.2.2, and 4.4.3, the assertion follows similarly for the bootstrap process. We note that for the density based bootstrap the covariance between X_t and X_s is even equal to zero, if the difference between t and s is large enough. □

Now, we can show the validity of the results of the previous sections and start with the results for the residual based bootstrap.

Proof of Lemma 4.2.1. By construction, $((b_t^*, e_t^*),\ t \in \mathbb{N})$ are i.i.d. sequences of random variables and b_i and e_j are independent for all $i \neq j$. Thus, the assertions about the moments of the bootstrap random variables are proven in Lemmas 3.7.1 through 3.7.3. For the logarithmic moment we obtain

$$\begin{aligned} 0 \leqslant \mathbf{E}^* \left[\ln^+ |e_0^*|\right] &= \frac{1}{N_e} \sum_{i=1}^{N_e} \max\left(\ln |\hat{e}_{T_i}|, 0\right) \\ &= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}} \sum_{t=1}^{n} \max\left(\ln |\hat{e}_t|, 0\right) \xrightarrow{\text{i.P.}} \mathbf{E}\left[\max\left(\ln |e_t|, 0\right)\right] < \infty, \end{aligned}$$

since

$$\begin{aligned} 0 \leqslant \max\left(\ln |\hat{e}_t|, 0\right) \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} &= \max\left(\ln |e_t - b_t X_{t-1} + (\hat{\varphi} - \varphi)\, X_{t-1}|, 0\right) \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} \\ &\leqslant \max\left(\ln |e_t|, 0\right) + \max\left(\ln(\,|e_t| + \mathcal{O}_P(\varepsilon)\,) - \ln |e_t|, 0\right) \\ &= \max\left(\ln |e_t|, 0\right) + o_P(1) \end{aligned}$$

and (c.f. Lemma 3.7.1)

$$\frac{1}{2n\varepsilon} \sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}} = f(0) + o_P(1).$$

A similar argumentation can be used to show the validity of the remaining statements noting that

$$|\hat{\varphi} + \hat{b}_t| = \left|\hat{\varphi} + \frac{X_t}{X_{t-1}} - \hat{\varphi}\right| = \left|\varphi + b_t + \frac{e_t}{X_{t-1}}\right| \leqslant |\varphi + b_t| + \left|\frac{e_t}{M}\right|.$$

The convergence of the bootstrap process, Equation (4.6), follows from Aue et al. (2006), Lemma 1, with the other results of the Lemma. □

Proof of Lemma 4.2.2. In the following, we denote by F_e the distribution function of the random variable e and by $\hat{F}_e$ its empirical distribution function based on independent observations $e_1, \ldots, e_n$. We then know from the Glivenco Cantelli Theorem (Klenke (2006), Satz 5.23) that

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{F}_e(x) - F_e(x)| = 0 \quad a.s. \tag{4.9}$$

so by Definition 4.11 from Elstrodt (2009) and the Theorem of Helly-Bray (Elstrodt (2009), Satz 4.13) for two random variables e_1 and e_2 :

$$\int \hat{F}_{e_1}(x-y) d\hat{F}_{e_2}(y) - \int \hat{F}_{e_1}(x-y) dF_{e_2}(y) \xrightarrow{n\to\infty} 0 \quad a.s.$$

and

$$\begin{aligned}
\hat{F}_{e_1+e_2}(x) - F_{e_1+e_2}(x) &= \int \hat{F}_{e_1}(x-y)d\hat{F}_{e_2}(y) - \int F_{e_1}(x-y)dF_{e_2}(y) \\
&= \int \hat{F}_{e_1}(x-y)d\hat{F}_{e_2}(y) - \int \hat{F}_{e_1}(x-y)dF_{e_2}(y) + \int \left(\hat{F}_{e_1}(x-y) - F_{e_1}(x-y)\right) dF_{e_2}(y) \\
&\overset{n\to\infty}{\longrightarrow} 0 \quad a.s..
\end{aligned}$$

For arbitrary r we recall the definition of $\tilde{X}_t^r$. If we set $F_{r,\varphi}(x) \equiv F_{\tilde{X}_t^r}(x)$ to be the distribution function of $\tilde{X}_t^r$ and $\hat{F}_{r,\varphi}(x)$ to be its empirical distribution function constructed as above from the empirical distribution functions of $(b_t)_{t\in\mathbb{Z}}$ and $(e_t)_{t\in\mathbb{Z}}$ we obtain per induction and a similar argumentation as before that $\hat{F}_{r,\varphi}(x) \overset{n\to\infty}{\longrightarrow} F_{r,\varphi}(x)$ $a.s.$ $\forall\, x \in \mathbb{R}$. Additionally, we know

$$\begin{aligned}
F_e^*(x) = P^*(e_0^* \leqslant x) &= \frac{1}{N_e}\sum_{i=1}^{N_e} \mathbb{1}_{\left\{X_{t_i} - \hat{\varphi}X_{t_i-1} \leqslant x\right\}} \\
&= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}} \sum_{t=1}^n \mathbb{1}_{\{e_t \leqslant x - b_t X_{t-1} - (\varphi-\hat{\varphi})X_{t-1}, |X_{t-1}|\leqslant\varepsilon\}} \\
&= \frac{1}{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}} \frac{1}{n}\sum_{t=1}^n \left(\mathbb{1}_{\{e_t\leqslant x\}} + \mathbb{1}_{\{e_t \leqslant x+R, |R|\leqslant\varepsilon(|b_t|+|\varphi-\hat{\varphi}|)\}} - \mathbb{1}_{\{e_t\leqslant x\}}\right)\mathbb{1}_{\{|X_{t-1}|<\varepsilon\}}
\end{aligned}$$

so that we have for $|\vartheta| \leqslant \varepsilon\left(|b_t| + |\varphi - \hat{\varphi}|\right)$

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{x\in\mathbb{R}} \left| F_e^*(x) - \hat{F}_e(x)\right| \\
&\quad\leqslant \limsup_{n\to\infty}\sup_{x\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^n \left(\mathbb{1}_{\{e_t\leqslant x+|\vartheta|\}} - \mathbb{1}_{\{e_t\leqslant x\}}\right) - F(x+|\vartheta|) + F(x) + F(x+|\vartheta|) - F(x) \\
&\quad= 0, \quad \text{because of Equation (4.9) and because } F \text{ is continuous from above.}
\end{aligned}$$

Hence, we obtain with the argumentation used before that $F^*_{\tilde{X}_t^{*,r}} - \hat{F}_{r,\hat{\varphi}} \overset{n\to\infty}{\longrightarrow} 0$ $i.P.$ $\forall\, x \in \mathbb{R}$. Further, we know that $\hat{F}_{\varphi+b}(x) = \hat{F}_b(x-\varphi)$ and obtain from the continuous mapping theorem that $\hat{F}_{\hat{\varphi}+b}(x) - \hat{F}_{\varphi+b}(x) \overset{n\to\infty}{\longrightarrow} 0$ $i.P.$, so that $\hat{F}_{r,\hat{\varphi}}(x) - \hat{F}_{r,\varphi}(x) \overset{n\to\infty}{\longrightarrow} 0$ $i.P.$ $\forall\, x \in \mathbb{R}$. Moreover, we know from Lemma 4.6.1 that for r sufficiently large $\left|F_{\tilde{X}_t^r}(x) - F_{X_t}(x)\right| \leqslant \varepsilon$ $a.s.$, and by the same argumentation that is used in the proof there, we obtain that $\left|F^*_{\tilde{X}_t^{*,r}}(x) - F^*_{X_t^*}(x)\right| \leqslant \varepsilon$ $a.s.$
Putting all previous results together we obain $\forall\, x \in \mathbb{R}$:

$$\begin{aligned}
|F^*_{X_t^*}(x) - F_{X_t}(x)| \leqslant\, & |F^*_{X_t^*}(x) - F^*_{\tilde{X}_t^{*,r}}(x)| + |F^*_{\tilde{X}_t^{*,r}}(x) - \hat{F}_{r,\hat{\varphi}}| \\
&+ |\hat{F}_{r,\hat{\varphi}}(x) - \hat{F}_{r,\varphi}(x) + \hat{F}_{r,\varphi}(x) - F_{\tilde{X}_t^r}(x)| + |F_{\tilde{X}_t^r}(x) - F_{X_t}(x)| \overset{\text{i.P.}}{\longrightarrow} 0.
\end{aligned}$$

This yields for bounded functions f by Definition 4.11 from Elstrodt (2009) and the Theorem of Helly-Bray (Satz 4.13, Elstrodt (2009)):

$$\mathbf{E}^*\left[f(X_t^*)\right] = \int f(x)dF^*_{X_t^*}(x) \overset{n\to\infty}{\longrightarrow} \int f(x)dF_{X_t}(x) = \mathbf{E}\left[f(X_t)\right] \quad i.P.$$

and hence the assertion follows. □

Proof of Lemma 4.2.3. With $Y^*_{i-1}(u^*,\kappa,\gamma) := \dfrac{X^{*\kappa}_{i-1}}{\left(x^2 {X^*_{i-1}}^2 + y\right)^\gamma}$ simple computations yield that $g'(X^*_i, u^*) =$

$$\begin{pmatrix} -2\left(\varphi^* - s^* + b^*_i\right) Y^*_{i-1}(u^*,2,1) - 2e^*_i Y^*_{i-1}(u^*,1,1) \\ Y^*_{i-1}(u^*,2,1) - \left(\varphi^* - s^* + b^*_i\right)^2 Y^*_{i-1}(u^*,4,2) - 2b^*_i e^*_i Y^*_{i-1}(u^*,3,2) - e^{*2}_i Y^*_{i-1}(u^*,2,2) \\ Y^*_{i-1}(u^*,0,1) - \left(\varphi^* - s^* + b^*_i\right)^2 Y^*_{i-1}(u^*,2,2) - 2b^*_i e^*_i Y^*_{i-1}(u^*,1,2) - e^{*2}_i Y^*_{i-1}(u^*,0,2) \end{pmatrix}$$

and further that

$$\begin{aligned}
A^*_{11} &= \mathbf{E}^*\left[g'_1(X^*_i,u^*)^2\right] = 4\hat\omega^2 \mathbf{E}^*\left[Y^*_0(\hat\theta,4,2)\right] + 4\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,2)\right], \\
A^*_{22} &= \mathbf{E}^*\left[g'_2(X^*_i,u^*)^2\right] \\
&= \mathbf{E}^*\left[b^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,8,4)\right] + 4\hat\omega^2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,6,4)\right] + \mathbf{E}^*\left[e^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,4,4)\right] \\
&\quad + \mathbf{E}^*\left[Y^*_0(\hat\theta,4,2)\right] + 2\hat\omega^2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,6,4)\right] - 2\hat\omega^2\mathbf{E}^*\left[Y^*_0(\hat\theta,6,3)\right] \\
&\quad - 2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,4,3)\right], \\
A^*_{33} &= \mathbf{E}^*\left[g'_3(X^*_i,u^*)^2\right] \\
&= \mathbf{E}^*\left[b^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,4,4)\right] + 4\hat\omega^2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,4)\right] + \mathbf{E}^*\left[e^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,0,4)\right] \\
&\quad + \mathbf{E}^*\left[Y^*_0(\hat\theta,0,2)\right] + 2\hat\omega^2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,4)\right] - 2\hat\omega^2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,3)\right] \\
&\quad - 2\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,0,3)\right], \\
A^*_{12} &= A^*_{21} = \mathbf{E}^*\left[g'_1(X^*_i,u^*)g'_2(X^*_i,u^*)\right] = 0, \\
A^*_{13} &= A^*_{31} = \mathbf{E}^*\left[g'_1(X^*_i,u^*)g'_3(X^*_i,u^*)\right] = 0, \\
A^*_{23} &= A^*_{32} = \mathbf{E}^*\left[g'_2(X^*_i,u^*)g'_3(X^*_i,u^*)\right] \\
&= \mathbf{E}^*\left[b^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,6,4)\right] + \mathbf{E}^*\left[e^{*4}_0\right]\mathbf{E}^*\left[Y^*_0(\hat\theta,2,4)\right] + 6\hat\omega^2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,4,4)\right] \\
&\quad + \mathbf{E}^*\left[Y^*_0(\hat\theta,2,2)\right] - 2\hat\omega^2\mathbf{E}^*\left[Y^*_0(\hat\theta,4,3)\right] - 2\hat\sigma^2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,3)\right],
\end{aligned}$$

which yields the existence of A^*, the non-singularity can be obtained promptly analogous to the proof of Lemma 6 in Aue et al. (2006) and the convergence follows directly with Lemma 4.4.3. Additional computations yield

$$\begin{aligned}
H^*_{11} &= \mathbf{E}\left[\frac{\partial^2}{\partial s^2} g(X^*_0,\hat\theta)\right] = 2\mathbf{E}^*\left[Y^*_0(\hat\theta,2,1)\right], \quad H^*_{12} = H^*_{21} = \mathbf{E}\left[\frac{\partial^2}{\partial s \partial x} g(X^*_0,\hat\theta)\right] = 0, \\
H^*_{22} &= \mathbf{E}\left[\frac{\partial^2}{\partial x^2} g(X^*_0,\hat\theta)\right] = \mathbf{E}^*\left[Y^*_0(\hat\theta,4,2)\right], \quad H^*_{33} = \mathbf{E}\left[\frac{\partial^2}{\partial y^2} g(X^*_0,\hat\theta)\right] = \mathbf{E}^*\left[Y^*_0(\hat\theta,0,2)\right], \\
H^*_{13} &= H^*_{31} = \mathbf{E}\left[\frac{\partial^2}{\partial s \partial y} g(X^*_0,\hat\theta)\right] = 0, \; H^*_{23} = H^*_{32} = \mathbf{E}\left[\frac{\partial^2}{\partial x \partial y} g(X^*_0,\hat\theta)\right] = \mathbf{E}^*\left[Y^*_0(\hat\theta,1,2)\right].
\end{aligned}$$

This directly yields the existence of H^*, whereas the non-singularity can be obtained by an argumentation similar to the one of the proof by Aue et al. (2006), Lemma 8, and the convergence follows directly by applying Lemma 4.4.3. With Lemma 4.6.2 the convergence of $\frac{1}{n}\sum_{i=1}^n g''(X^*_i,\hat\theta)$ follows directly. □

Proof of Lemma 4.2.4. Let $C^* \subset \Gamma^*$ be a compact set such that the distance between $\hat{\theta}$ and C is positive. Using Lemma 4.6.2 we have

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{u^*\in C^*} \left(\frac{1}{n}\sum_{i=1}^n g(X_i^*, u^*) - \frac{1}{n}\sum_{i=1}^n g(X_i, u^*) \right) \\
&\quad = \liminf_{n\to\infty} \inf_{u^*\in C^*} \big(\mathbf{E}\left[g(X_0,u^*)\right] + o_P(1) - \mathbf{E}\left[g(X_0,u^*)\right] + o_P(1)\big) = 0 \quad \text{i. } P^*\text{-Pr. i. } P\text{-Pr.}
\end{aligned}$$

yielding

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{u^*\in C^*} l_n(X^*, u^*) - \mathbf{E}\left[g(X_0,\theta)\right] \\
&\quad = \liminf_{n\to\infty} \inf_{u^*\in C^*} \left(\frac{1}{n}\sum_{i=1}^n g(X_i^*, u^*) - \frac{1}{n}\sum_{i=1}^n g(X_i, u^*) + \frac{1}{n}\sum_{i=1}^n g(X_i, u^*) \right) - \mathbf{E}\left[g(X_0,\theta)\right] \\
&\quad \geqslant \liminf_{n\to\infty} \inf_{u^*\in C^*} l_n(X, u^*) + \liminf_{n\to\infty} \inf_{u^*\in C^*} \left(\frac{1}{n}\sum_{i=1}^n g(X_i^*, u^*) - \frac{1}{n}\sum_{i=1}^n g(X_i, u^*) \right) - \mathbf{E}\left[g(X_0,\theta)\right] \\
&\quad = \liminf_{n\to\infty} \inf_{u^*\in C^*} l_n(X, u^*) - \mathbf{E}\left[g(X_0,\theta)\right] \quad \text{i. } P^*\text{-Pr. i. } P\text{-Pr.}
\end{aligned}$$

The last term is strictly positive according to Aue et al. (2006), Equation (37). Since we know that $\hat{\theta}_n \in \Gamma^*$ we have further that $\inf_{u^*\in\Gamma^*} l_n(X^*, u^*) \leqslant l_n(X^*, \hat{\theta}_n)$ and thus

$$\begin{aligned}
\limsup_{n\to\infty} l_n(X^*, \hat{\theta}_n^*) = \limsup_{n\to\infty} \inf_{u^*\in\Gamma^*} l_n(X^*, u^*) &\leqslant \limsup_{n\to\infty} l_n(X^*, \hat{\theta}_n) \\
&= \lim_{n\to\infty} \mathbf{E}^* \left[g(X_0^*, \hat{\theta}_n)\right] + o_{P^*}(1)
\end{aligned}$$

so that we obtain $\limsup_{n\to\infty} l_n(X^*, \hat{\theta}_n^*) \leqslant \mathbf{E}\left[g(X_0,\theta)\right]$ i. P^*-Pr. i. P-Pr. since $\hat{\theta} \xrightarrow{\text{i.P.}} \theta$. Let $U \subset \Gamma^* \subset \Gamma$ be an open ball around θ with small enough radius. We know from Aue et al. (2006) that there exists an m_0 so that for all $n > m_0$ $\hat{\theta}_n \in U$. If $\hat{\theta}_n^* \notin U$ infinitely often for $n > n_0 > m_0$ there is a random subsequence n_k such that we have for $C^* = \Gamma^* \backslash U$

$$\begin{aligned}
\mathbf{E}\left[g(X_0,\theta)\right] < \liminf_{n\to\infty} \inf_{u^*\in C^*} l_n(X^*, u^*) &\leqslant \liminf_{k\to\infty} l_{n_k}(X^*, \hat{\theta}_{n_k}^*) \\
&\leqslant \limsup_{n\to\infty} l_n(X^*, \hat{\theta}_n^*) \leqslant \mathbf{E}\left[g_1(\theta)\right] \quad \text{i. } P^*\text{-Pr. i. } P\text{-Pr.}
\end{aligned}$$

This is a contradiction, such that there has to be an n_0 such that $\hat{\theta}_n^* \in U^*$ $\forall\, n > n_0$, i.e. $\hat{\theta}_n^* - \hat{\theta}_n \xrightarrow{n\to\infty} 0$ i. P^*-Pr. i. P-Pr.. □

Proof of Lemma 4.2.5. To show the assertion, we use the Cramer Wold device (see Billingsley (1968), Theorem 7.7) and a central limit theorem for weak dependent random variables (see Neumann & Paparoditis (2008), Theorem 6.1). According to the Cramer Wold device the assertion of the Lemma holds if and only if

$$c^\top \sqrt{n}\, l_n'(X^*, \hat{\theta}) \xrightarrow{\mathrm{D}} \mathcal{N}(0, c^\top A c) \quad i.P. \quad \forall\, c \in \mathbb{R}^3.$$

We set $Y_{i-1}^*(\hat{\theta}, \kappa, \gamma)$ as above and

$$Z_{n,i}^* = \begin{pmatrix} -2b_i^* Y_{i-1}^*(\hat{\theta},2,1) - 2e_i^* Y_{i-1}^*(\hat{\theta},1,1) \\ -b_i^{*2} Y_{i-1}^*(\hat{\theta},4,2) - 2b_i^* e_i^* Y_{i-1}^*(\hat{\theta},3,2) - e_i^{*2} Y_{i-1}^*(\hat{\theta},2,2) + Y_{i-1}^*(\hat{\theta},2,1) \\ -b_i^{*2} Y_{i-1}^*(\hat{\theta},2,2) - 2b_i^* e_i^* Y_{i-1}^*(\hat{\theta},1,2) - e_i^{*2} Y_{i-1}^*(\hat{\theta},0,2) + Y_{i-1}^*(\hat{\theta},0,1) \end{pmatrix}$$

and consider for arbitrary $c \in \mathbb{R}^3$

$$c^\top \sqrt{n}\, l_n'(X^*, \hat{\theta}) = \sqrt{n} \sum_{i=1}^{n} c^\top Z_{n,i}^* = \sqrt{n} \sum_{i=1}^{n} W_{n,i}^*$$

with $\mathbb{R}$-valued $W_{n,i}^* = c^\top Z_{n,i}^*,\ i = 1, \ldots, n$. The aforementioned theorem now can be applied on $\left(W_{n,i}^*\right)_{i=1,\ldots,n}$ if we can show that the prerequisites are met:

Autocovariances: $\mathbf{Cov}^*\left[W_{n,i}^*, W_{n,j}^*\right] = \mathbf{Cov}^*\left[c^\top Z_{n,i}^*, c^\top Z_{n,j}^*\right]$ consists of terms of the form $\mathbf{Cov}^*\left[{Z_{n,i}^*}^{(a)}, {Z_{n,j}^*}^{(b)}\right],\ a, b = 1, 2, 3$, and by Lemma 4.4.3 these terms converge in probability to $\mathbf{Cov}\left[{Z_{n,i}}^{(a)}, {Z_{n,j}}^{(b)}\right],\ a, b = 1, 2, 3$. Aue et al. (2006) showed that these terms are equal to zero for $i \neq j$ (see the proof of Lemma 7 therein), hence $W_{n,i}^*$ and $W_{n,j}^*$ are asymptotically uncorrelated in probability for $i \neq j$. Since

$$\mathbf{E}^*\left[W_{n,0}^{*^2}\right] = \mathbf{E}^*\left[\left(c^\top Z_{n,0}^*\right)^2\right] = \mathbf{E}^*\left[c^\top Z_{n,0}^* Z_{n,0}^{*^\top} c\right] = c^\top \mathbf{E}^*\left[Z_{n,0}^* Z_{n,0}^{*^\top}\right] c = c^\top A^* c$$

the convergence of the autocovariances is now straightforward:

$$\sigma_n^{*2} = \mathbf{E}^*\left[W_{n,0}^{*^2}\right] + 2 \sum_{k=1}^{n} \mathbf{E}^*\left[W_{n,0}^* W_{n,k}^*\right] \xrightarrow{n \to \infty} c^\top A c \quad i.P.$$

Lindeberg condition: Let $\varepsilon > 0$.

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}^*\left[W_{n,k}^{*^2} \mathbb{1}_{|W_{n,k}^*| \geqslant \varepsilon\sqrt{n}}\right] &\leqslant \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}^*\left[|W_{n,k}^*|^3\right]^{\frac{2}{3}} \mathbf{E}^*\left[\mathbb{1}_{|W_{n,k}^*| \geqslant \varepsilon\sqrt{n}}\right]^{\frac{1}{3}} \\ &= \sum_{k=1}^{n} \mathbf{E}^*\left[|W_{n,1}^*|^3\right]^{\frac{2}{3}} P^*\left\{|W_{n,1}^*| \geqslant \varepsilon\sqrt{n}\right\}^{\frac{1}{3}} \\ &\leqslant \frac{1}{\varepsilon \sqrt[3]{n}} \mathbf{E}^*\left[|W_{n,1}^*|^3\right]^{\frac{2}{3}} \mathbf{E}^*\left[W_{n,1}^{*^2}\right]^{\frac{1}{3}} \xrightarrow{n \to \infty} 0 \quad i.P., \end{aligned}$$

if $\mathbf{E}^*\left[|W_{n,1}^*|^3\right] = \mathcal{O}_P(1)$, i.e. $\mathbf{E}^*\left[b_t^{*^6}\right], \mathbf{E}^*\left[e_t^{*^6}\right] = \mathcal{O}_P(1)$.

Weak-dependance conditions: Let f be a measurable and square-integrable function and $s_1 < \cdots < s_u < s_u + v = t_1$ and $v \geqslant 1$. Then,

$$\begin{aligned} &\mathbf{Cov}^*\left[f(W_{n,s_1}^*, \ldots, W_{n,s_u}^*), W_{n,t_1}^*\right] \\ &\qquad = \mathbf{E}^*\left[f(W_{n,s_1}^*, \ldots, W_{n,s_u}^*)\left(c_1 Z_{n,t_1}^{*^{(1)}} + c_2 Z_{n,t_1}^{*^{(2)}} + c_3 Z_{n,t_1}^{*^{(3)}}\right)\right] \\ &\qquad\quad - \mathbf{E}^*\left[f(W_{n,s_1}^*, \ldots, W_{n,s_u}^*)\right] \mathbf{E}^*\left[c_1 Z_{n,t_1}^{*^{(1)}} + c_2 Z_{n,t_1}^{*^{(2)}} + c_3 Z_{n,t_1}^{*^{(3)}}\right] \end{aligned}$$

Now, we consider the term that contains $Z_{n,t_1}^{*^{(1)}} = -2 b_{t_1}^* Y_{t_1-1}^*(\hat{\theta}, 2, 1) - 2 e_{t_1}^* Y_{t_1-1}^*(\hat{\theta}, 1, 1)$. Since $b_{t_1}^*$ and $e_{t_1}^*$ are independent of all other terms, we can separate these terms from each expectation, and since both have zero mean, the respective terms vanish. Next, we consider the term that contains $Z^{*(2)}_{n,t_1} = -b^{*2}_{t_1} Y^*_{t_1-1}(\hat\theta,4,2) - 2b^*_{t_1} e^*_{t_1} Y^*_{t_1-1}(\hat\theta,3,2) - e^{*2}_{t_1} Y^*_{t_1-1}(\hat\theta,2,2) + Y^*_{t_1-1}(\hat\theta,2,1)$. Again, we can separate the terms $b^{*1,2}_{t_1}$ and $e^{*1,2}_{t_1}$. Regarding that both have zero mean and second moment ω^2 or σ^2, respectively, and merging everything again, we obtain

$$\begin{aligned}
&\mathbf{E}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u})c_2 Z^{*(2)}_{n,t_1}\right]\\
&\quad= c_2\mathbf{E}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u})\left(-\hat\omega^2 Y^*_{i-1}(\hat\theta,4,2) - \hat\sigma^2 Y^*_{i-1}(\hat\theta,2,2) + Y^*_{t_1-1}(\hat\theta,2,1)\right)\right]\\
&\quad= c_2\mathbf{E}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u})\left(-\frac{\hat\omega^2 X^{*4}_{t_1-1} + \hat\sigma^2 X^{*2}_{t_1-1}}{\left(\hat\omega^2 X^{*2}_{t_1-1} + \hat\sigma^2\right)^2} + \frac{X^{*2}_{t_1-1}\left(\hat\omega^2 X^{*2}_{t_1-1} + \hat\sigma^2\right)}{\left(\hat\omega^2 X^{*2}_{t_1-1} + \hat\sigma^2\right)^2}\right)\right] = 0,
\end{aligned}$$

and in the same way $\mathbf{E}^*\left[Z^{*(2)}_{n,t_1}\right] = 0$ as well. By similar computations, it can be seen that $\mathbf{E}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u})c_2 Z^{*(3)}_{n,t_1}\right] = 0 = \mathbf{E}^*\left[Z^{*(3)}_{n,t_1}\right]$ as well so that we finally obtain

$$\mathbf{Cov}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u}), W^*_{n,t_1}\right] = 0.$$

Let now f be measurable and bounded and $s_1 < \dots < s_u < s_u + v = t_1 \leqslant t_2$ and $v \geqslant 1$.

$$\begin{aligned}
&\mathbf{Cov}^*\left[f(W^*_{n,s_1},\dots,W^*_{n,s_u}), W^*_{n,t_1}\cdot W^*_{n,t_2}\right]\\
&\quad= \mathbf{Cov}^*\left[[f(W^*_{n,s_1},\dots,W^*_{n,s_u}), c^\top Z^*_{n,t_1}\cdot c^\top Z^*_{n,t_2}\right]\\
&\quad= \mathbf{Cov}^*\Big[[f(W^*_{n,s_1},\dots,W^*_{n,s_u}), c_1^2 Z^{*(1)}_{n,t_1} Z^{*(1)}_{n,t_2} + c_2^2 Z^{*(2)}_{n,t_1} Z^{*(2)}_{n,t_2} + c_3^2 Z^{*(3)}_{n,t_1} Z^{*(3)}_{n,t_2} + c_1c_2 Z^{*(1)}_{n,t_1} Z^{*(2)}_{n,t_2}\\
&\qquad\quad + c_1c_3 Z^{*(1)}_{n,t_1} Z^{*(3)}_{n,t_2} + c_2c_3 Z^{*(2)}_{n,t_1} Z^{*(3)}_{n,t_2} + c_1c_2 Z^{*(2)}_{n,t_1} Z^{*(1)}_{n,t_2} + c_1c_3 Z^{*(3)}_{n,t_1} Z^{*(1)}_{n,t_2} + c_2c_3 Z^{*(3)}_{n,t_1} Z^{*(2)}_{n,t_2}\Big]
\end{aligned}$$

We split this term up and consider just the first term that is equal to

$$\begin{aligned}
\mathbf{Cov}^*\Big[f(W^*_{n,s_1},...,W^*_{n,s_u}),&\Big(b^*_{t_1} Y^*_{t_1-1}(\hat\theta,2,1) b^*_{t_2} Y^*_{t_2-1}(\hat\theta,2,1) + b^*_{t_1} Y^*_{t_1-1}(\hat\theta,2,1) e^*_{t_2} Y^*_{t_2-1}(\hat\theta,1,1)\\
&+ e^*_{t_1} Y^*_{t_1-1}(\hat\theta,1,1) b^*_{t_2} Y^*_{t_2-1}(\hat\theta,2,1) + e^*_{t_1} Y^*_{t_1-1}(\hat\theta,1,1) e^*_{t_2} Y^*_{t_2-1}(\hat\theta,1,1)\Big)\Big].
\end{aligned}$$

Again, we can write this term as expectations and can separate either $b^*_{t_{1,2}}$ or $e^*_{t_{1,2}}$ from each term, so that this term is equal to zero. For the second term we obtain

$$\begin{aligned}
\mathbf{Cov}^*\Big[f(W^*_{s_1},...,W^*_{s_u}),&\Big(b^{*2}_{t_1} Y^*_{t_1-1}(\hat\theta,4,2)\left(b^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,4,2) + e^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,2,2) - Y^*_{t_2-1}(\hat\theta,1,1)\right)\\
&+ 2b^*_{t_1} e^*_{t_1} Y^*_{t_1-1}(\hat\theta,3,2)\left(b^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,4,2) + e^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,2,2) - Y^*_{t_2-1}(\hat\theta,1,1)\right)\\
&+ e^{*2}_{t_1} Y^*_{t_1-1}(\hat\theta,2,2)\left(b^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,4,2) + e^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,2,2) - Y^*_{t_2-1}(\hat\theta,1,1)\right)\\
&- Y^*_{t_1-1}(\hat\theta,2,1)\left(b^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,4,2) + e^{*2}_{t_2} Y^*_{t_2-1}(\hat\theta,2,2) - Y^*_{t_2-1}(\hat\theta,1,1)\right)\Big)\Big]
\end{aligned}$$

Writing this as an expectation gives by a similar argumentation as above that the first summand

$$\mathbf{E}^*\left[f(W^*_{n,s_1},...,W^*_{n,s_u}) b^{*2}_{t_1} Y^*_{t_1-1}(\hat\theta,4,2)\left(\omega^2 Y^*_{t_2-1}(\hat\theta,4,2) + \sigma^2 Y^*_{t_2-1}(\hat\theta,2,2) - Y^*_{t_2-1}(\hat\theta,1,1)\right)\right]$$

is equal to zero. Similarly, all other term vanish as well, so that we obtain

$$\mathbf{Cov}^* \left[f(W^*_{n,s_1}, \ldots, W^*_{n,s_u}), W^*_{n,t_1} \cdot W^*_{n,t_2} \right] = 0.$$

Hence, the aforementioned theorem can be applied to obtain the desired result. □

Now, we consider the results we have for the density based bootstrap and first have another result concerning the density estimation.

Lemma 4.6.3. *Let $X_1^{(1)}, \ldots, X_n^{(1)}$ and $X_1^{(2)}, \ldots, X_n^{(2)}$ each be i.i.d. random variables with density k_1 and k_2, respectively, and $\hat{k}_{1,n}$ and $\hat{k}_{2,n}$ the kernel density estimators of the form*

$$\hat{k}_{n,1}(x) = \hat{k}_{n,2}(x) = \frac{1}{n} \sum_{i=1}^{n} K(X_i - x)$$

with a kernel function K for k_1 and k_2 such that $\sup_{x \in [-2R,2R]} \left| \hat{k}_{n,1}(x) - k_1(x) \right| = \mathcal{O}(a_n)$ and $\sup_{x \in [-2R,2R]} \left| \hat{k}_{n,2}(x) - k_2(x) \right| = \mathcal{O}(a_n)$ for a sequence $(a_n)_{n \in \mathbb{N}} \overset{n\to\infty}{\longrightarrow} 0$. With $R = R(n) \overset{n\to\infty}{\longrightarrow} \infty$ such that $Ra_n \overset{n\to\infty}{\longrightarrow} 0$, $\forall\, x \in [-R, R]$ the estimator

$$\hat{k}_n^{(12)}(x) = \int_{-R}^{R} \hat{k}_{n,1}(u) \hat{k}_{n,2}(x-u)\, du = k_n^{(12)}(x) + \mathcal{O}_P\left(R a_n \right)$$

is consistent for the density $k^{(12)}(x) = k_1 * k_2(x) = \int k_1(u) k_2(x-u) du$ of the convolution $X_1^{(1)} + X_1^{(2)}$.

Proof. We consider for $x \in [-R, R]$

$$\begin{aligned}
&\mathbf{E}\left[\hat{k}_1(u) \hat{k}_2(x-u) \right] \\
&\qquad = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{E}\left[K\left(X_i - u \right) K\left(X_j - x + u \right) \right] \\
&\qquad = \frac{n-1}{n} \mathbf{E}\left[K\left(X_1 - u \right) \right] \mathbf{E}\left[K\left(X_1 - x + u \right) \right] + \frac{1}{n} \mathbf{E}\left[K\left(X_1 - u \right) K\left(X_1 - x + u \right) \right] \\
&\qquad = \frac{n-1}{n} \left(k_1(u) + \mathcal{O}(a_n) \right) \left(k_2(x-u) + \mathcal{O}(a_n) \right) + \mathcal{O}(a_n) \\
&\qquad = k_1(u) k_2(x-u) + \mathcal{O}\left(a_n \right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{Var}\left[\hat{k}_1(u) \hat{k}_2(x-u) \right] \\
&\qquad = \frac{1}{n^4} \mathbf{Var}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} K\left(X_i - u \right) K\left(X_j - x + u \right) \right] \\
&\qquad \leqslant \frac{1}{n^4} \mathbf{E}\left[\left(\sum_{i=1}^{n} \sum_{j=1}^{n} K\left(X_i - u \right) K\left(X_j - x + u \right) \right)^2 \right] \\
&\qquad = \frac{1}{n^4} \mathbf{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} K\left(X_i - u \right) K\left(X_k - u \right) K\left(X_j - x + u \right) K\left(X_l - x + u \right) \right] \\
&\qquad = \mathcal{O}\left(a_n \right),
\end{aligned}$$

where the sequence $(a_n)_{n\in\mathbb{N}}$ can be chosen independently of x and u. Hence,

$$\sup_{x\in[-R,R]}\left|\hat{k}_{n,1}(u)\hat{k}_{n,2}(x-u)-k_1(u)k_2(x-u)\right|=\mathcal{O}_P(a_n)$$

so that integration yields the desired result. $\square$

Proof of Lemma 4.4.1. By easy transformations it can be seen that the process $\tilde{X}_t^r$ can be written as

$$\tilde{X}_t^r=(\varphi+b_t)\left((\varphi+b_{t-1})\left((\varphi+b_{t-2})\left(\cdots\left((\varphi+b_{t-r+2})\,e_{t-r+1}\right)+\cdots\right)+e_{t-2}\right)+e_{t-1}\right)+e_t.$$

Hence, the density f_r of $\tilde{X}_t^r$ is the convolution of the densities of the respective random variables $e_{t-r},\dots,e_t$ and $b_{t-r+1},\dots,b_t$. Since the sequences e_t and b_t are each i.i.d. we can write

$$f_r(x)=\prod_{i=1}^{r}\prod_{j=1}^{r-1}\int\cdots\int h(x_i-\varphi)k\left(\frac{y_r}{x_r}\right)k\left(\frac{y_j}{x_j}-x_{j+1}\right)k(x-y_1)\mathrm{dx}_1\dots\mathrm{dx}_r\mathrm{dy}_1\dots\mathrm{dy}_r.$$

With the result of Lemma 4.6.3 we obtain per induction the desired result for $\hat{f}_{n,r}$. Note that for each integration we add one R to the Bias term and that the domain where x is allowed to be in shrinks with each integration with R on both sides. $\square$

Proof of Theorem 4.4.2. We know from Lemma 4.6.1 that for all $\tilde{\varepsilon}>0$ it exists a r sufficiently large such that $\forall\, r>r_0$: $\left|F_{\tilde{X}_t^r}(x)-F_{X_t}(x)\right|\leqslant\tilde{\varepsilon}$ *a.s.*, so that $\forall\,\varepsilon>0\;\exists\, r_0'$ such that $\forall\, r>r_0'$: $|f_r(x)-f(x)|<\varepsilon$. From Lemma 4.4.1 we conclude that we can find an n_0 for each r and arbitrary $\delta',\varepsilon'>0$ such that $\forall\, n>n_0$: $P\left\{\left|\hat{f}_{n,r}(x)-f_r(x)\right|>\delta\right\}<\varepsilon$ so that finally $\forall\,\delta,\varepsilon>0\;\exists\, m_0,s_0$ such that $\forall\, n>m_0$, $\forall\, r>s_0$:

$$P\left\{\left|\hat{f}_{n,r}(x)-f(x)\right|>\delta\right\}\leqslant P\left\{\left|\hat{f}_{n,r}(x)-f_r(x)\right|>\delta\right\}+P\left\{|f_r(x)-f(x)|>\delta\right\}\leqslant 2\varepsilon.$$

This yields the assertion. $\square$

Proof of Lemma 4.4.3. By construction, $((b_t^*,e_t^*),\ t\in\mathbb{N})$ are i.i.d. sequences of random variables and b_i and e_j are independent for all $i\neq j$, thus, the first two statements are consequences of Lemma 4.3.4 and the assumptions. Further, with $C_b<\infty$,

$$\begin{aligned}
0&\leqslant\mathbf{E}^*\left[\ln^+|\hat{\varphi}+b_0^*|\right]\\
&=\int_{-R}^{R}\ln|\hat{\varphi}+b|\mathbb{1}_{\{|\hat{\varphi}+b|\geqslant1\}}\hat{h}(b)\mathrm{db}\\
&=\int_{-R}^{R}\ln^+|\varphi+b|\hat{h}(b)\mathrm{db}+\int_{-R}^{R}\left(\ln|\hat{\varphi}+b|-\ln|\varphi+b|\right)\mathbb{1}_{\{|\hat{\varphi}+b|\geqslant1\}}\hat{h}(b)\mathrm{db}\\
&=\int_{-R}^{R}\ln^+|\varphi+b|\hat{h}(b)\mathrm{db}+\int_{-R}^{R}\left(\ln|\varphi+b|+\mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)C_b-\ln|\varphi+b|\right)\mathbb{1}_{\{|\hat{\varphi}+b|\geqslant1\}}\hat{h}(b)\mathrm{db}\\
&=\int_{-R}^{R}\ln^+|\varphi+b|\hat{h}(b)\mathrm{db}+\mathcal{O}_P\left(\frac{R}{\sqrt{n}}\right)\\
&\xrightarrow{\text{i.P.}}\mathbf{E}\left[\ln^+[|\varphi+b_0|\right]<\infty
\end{aligned}$$

by Definition 4.3.3 and a taylor expansion. The other two assertions follow by similar argumentations. The convergence and stationarity of X_t^* now follows by Aue et al. (2006), Theorem 1. The existence of the density as given in Lemma 4.4.1 follows directly from the strict stationarity and the fact that both the noises have densitites. Further,

$$\mathbf{E}^*\left[X_t^*\right] = \mathbf{E}^*\left[\sum_{i=0}^{r-1}\left(\prod_{j=0}^{i-1}(\hat{\varphi} + b_{t-j}^*)\right) e_{t-i}^*\right] = \sum_{i=0}^{r-1}\left(\prod_{j=0}^{i-1}(\hat{\varphi} + \mathbf{E}^*\left[b_{t-j}^*\right])\right)\mathbf{E}^*\left[e_{t-i}^*\right] = 0,$$

$$\mathbf{E}^*\left[X_t^{*^2}\right] = \sum_{i=0}^{r-1}\left(\prod_{j=0}^{i-1}\left(\hat{\varphi}^2 + \hat{\omega}^2\right)\right)\hat{\sigma}^2 = \hat{\sigma}^2\frac{1-(\hat{\varphi}^2+\hat{\omega}^2)^r}{1-\hat{\varphi}-\hat{\omega}^2} \xrightarrow{\text{i.P.}} \frac{\sigma^2}{1-\varphi-\omega^2} = \mathbf{E}\left[X_t^2\right].$$

Finally, $\exists\, n_0$ such that $\forall\, n > n_0 : \hat{\omega}_n^2 x^2 + \hat{\sigma}_n \in \omega^2 x^2 + \sigma^2 \pm (\varepsilon_1 x^2 + \varepsilon_2)$ since $\hat{\omega}_n^2 \xrightarrow{\text{a.s.}} \omega^2$ and $\hat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \sigma^2$. Hence, it exists a function $g(x)$ such that $g_n(x) = \frac{x^\kappa}{(\hat{\omega}^2x^2+\hat{\sigma}^2)^\gamma} \leqslant g(x)$ and $\int_{-R}^{R} g(x)\hat{f}_{n,r}(x)\mathrm{dx} < \infty$. Thus, by applying the continuous mapping theorem and dominated convergence and the way of convergence of $\hat{f}_{n,r}$ towards f such that the integral

$$\int\left(\frac{x^\kappa}{(\hat{\omega}^2x^2+\hat{\sigma}^2)^\gamma} - \frac{x^\kappa}{(\omega^2x^2+\sigma^2)^\gamma}\right)\hat{f}_{n,r}(x)\mathrm{dx} = o_P(1),$$

converges in probability as well and using that X_t has the density f and X_t^* the density $\hat{f}_{n,r}$ all this yields that

$$\begin{aligned}
&\mathbf{E}^*\left[\frac{X_{i-1}^{*^\kappa}}{\left(\omega^2 X_{i-1}^{*^2}+\sigma^2\right)^\gamma}\right] \\
&\quad = \int\frac{x^\kappa}{(\omega^2x^2+\sigma^2)^\gamma}\hat{f}_{n,r}(x)\mathrm{dx} + \int\left(\frac{x^\kappa}{(\hat{\omega}^2x^2+\hat{\sigma}^2)^\gamma} - \frac{x^\kappa}{(\omega^2x^2+\sigma^2)^\gamma}\right)\hat{f}_{n,r}(x)\mathrm{dx} \\
&\quad \xrightarrow{\text{i.P.}} \int\frac{x^\kappa}{(\omega^2x^2+\sigma^2)^\gamma}f(x)\mathrm{dx} = \mathbf{E}\left[\frac{X_{i-1}^\kappa}{\left(\omega^2X_{i-1}^2+\sigma^2\right)^\gamma}\right],
\end{aligned}$$

what concludes the proof of the Lemma. □

5 Estimation of densities of random coefficient autoregressive processes

In the previous chapter, we made use of estimators for the densities of both the innovation and the disturbance noise. For the bootstrap approaches mentioned there, any estimator fulfilling the requirements of Definition 4.3.3 can be used. However, we now want to thoroughly introduce and examine the estimators that we gave in Definitions 4.3.1 and 4.3.2.
Nonparametric estimation of densities of random variables is a well known field in statistics. Since the first contributions by Rosenblatt (1956) for independent and identically distributed random variables, a large amount of papers emerged. These also considered the problem of estimating the densities of the innovations in standard autoregression models, what is a well known tool today. More complicated is the density estimation of two convoluted random variables. Standard methods in this case, like the ones introduced by Fan (1991), amongst others, require the density of one random variable to be known if estimating the density of the other random variable. Following these two approaches for the innovation and the disturbance noise, we introduce kernel density estimators for the innovation noise of a random coefficient autoregressive process as well as for the disturbance noise of such a process. After having stated some estimators in the previous section already, we now thoroughly introduce these estimators and analyze their asymptotic behavior. First, we introduce an estimator for the innovation noise.

5.1 Innovation parameter

5.1.1 Derivation of the estimator

To let us start, we assume that we have an estimator $\hat{\varphi}$ of φ and consider the estimated residuals $\hat{u}_t^n = X_t - \hat{\varphi}X_{t-1}$. Following the idea that was used in the previous chapters to construct a bootstrap method, we use the "small" observations $|X_{t-1}| < \varepsilon$ to estimate the density of the innovation noise via a kernel density estimator $\hat{k}(\cdot)$ with a kernel K and bandwidth $h = h(n)$:

$$\hat{k}_{n,\varepsilon}(y) = \frac{1}{\sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}|\leqslant\varepsilon(n)\}}} \sum_{t=1}^{n} \frac{1}{h} K\left(\frac{\hat{u}_t^{n,\varepsilon} - y}{h}\right) \mathbb{1}_{\{|X_{t-1}|\leqslant\varepsilon(n)\}}$$

Clearly, this function is not continuous in X_{t-1}. Instead of using the indicator function, we could introduce another kernel function, W, with bandwidth $\varepsilon = \varepsilon(n)$. If we define this function independent of y it ensures that only these $\hat{u}_t^{n,\varepsilon}$ that correspond to a sufficient small value of $|X_{t-1}|$ are under consideration for the estimation of the density. Thus, we obtain the following estimator:

Definition 5.1.1. *In an RCA model, with $\hat{u}_t^n = X_t - \hat{\varphi}X_{t-1}$, an estimator for the density of the innovation noise e_t is given by*

$$\hat{k}_{n,\varepsilon}(y) = \frac{1}{\sum_{t=1}^n \frac{1}{\varepsilon} W\left(\frac{X_{t-1}}{\varepsilon}\right)} \sum_{t=1}^n \frac{1}{h\varepsilon} K\left(\frac{\hat{u}_t^n - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)$$

where K and W are two kernel functions and the bandwidths are $h = h(n) \stackrel{n\to\infty}{\longrightarrow} 0$ and $\varepsilon = \varepsilon(n) \stackrel{n\to\infty}{\longrightarrow} 0$.

Using the definition of $\hat{u}_t^n$ and a taylor expansion for K, we can write this estimator with $\vartheta \in [0,1]$ in a different way:

$$\begin{aligned}
\hat{k}(y) = & \frac{1}{\frac{1}{n\varepsilon}\sum_{t=1}^n W\left(\frac{X_{t-1}}{\varepsilon}\right)} \cdot \frac{1}{nh\varepsilon} \sum_{t=1}^n K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \\
& + \frac{1}{\frac{1}{n\varepsilon}\sum_{t=1}^n W\left(\frac{X_{t-1}}{\varepsilon}\right)} \cdot \frac{\varphi - \hat{\varphi}}{h} \frac{1}{nh\varepsilon} \sum_{t=1}^n X_{t-1} K'\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \\
& + \frac{1}{\frac{1}{n\varepsilon}\sum_{t=1}^n W\left(\frac{X_{t-1}}{\varepsilon}\right)} \cdot \left(\frac{\varphi - \hat{\varphi}}{h}\right)^2 \\
& \qquad \frac{1}{nh\varepsilon} \sum_{t=1}^n X_{t-1}^2 K''\left(\frac{b_t X_{t-1} + e_t - y}{h} + \vartheta \frac{\varphi - \hat{\varphi}}{h} X_{t-1}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \qquad (5.1) \\
:= & \frac{1}{A_2} \cdot A_1 + \frac{1}{A_2} \cdot B_1 + \frac{1}{A_2} \cdot B_2
\end{aligned}$$

We will use this decomposition to further analyze the estimator, but first we would like to state the conditions that we impose on the Kernel functions and the bandwidths for this analysis. These conditions are assumed to hold true for the remainder of the chapter.

Conditions 5.1.2.

- We consider two non-negative, symmetric and centered kernel functions K and W that are two times continuously differentiable. The Kernels and their first derivative are Lipschitz continuous. The second derivative of K is assumed to be bounded and we denote $\sup_x |K''(x)| = K''(\Xi'') < \infty$.
- The bandwidths satisfy $h = h(n) \stackrel{n\to\infty}{\longrightarrow} 0$, $\varepsilon = \varepsilon(n) \stackrel{n\to\infty}{\longrightarrow} 0$, $nh\varepsilon \stackrel{n\to\infty}{\longrightarrow} \infty$, $nh^2 \stackrel{n\to\infty}{\longrightarrow} \infty$, $\frac{\varepsilon^{1.5}}{n^{1.5}h^3} \stackrel{n\to\infty}{\longrightarrow} 0$, $nh^5\varepsilon + nh\varepsilon^5 \stackrel{n\to\infty}{\longrightarrow} c < \infty$, $\int K^i(u)|u^j|\mathrm{du} < \infty$, $(i = 1,2,\ j = 0,1,2,$ and $j = 0$ for $i = 3,4)$ $\int |K'(u)|\ |u^j|\mathrm{du} < \infty$, $(j = 0,1,2)$ $\int W^i(u)\mathrm{du} < \infty$, $(i = 3,4)$ $\int W^i(u)|u^j|\mathrm{du} < \infty$, $(i = 1,2,\ j = 0,1,2,3,4,5)$
- The density of X_t exists, is denoted by $f(\cdot)$ and is two times continuously differentiable with bounded second derivative: $\sup_x |f''(x)| = f''(\Gamma'') < \infty$. In addition, f is continuous in a neighborhood of 0 and $f(0) > 0$.
- $(e_t)_{t\in\mathbb{Z}}$ and $(b_t)_{t\in\mathbb{Z}}$ have finite fourth moments and are independent white noises.
- The density of e_t exists, is denoted by $k(\cdot)$ and is two times continuously differentiable with bounded second derivative: $\sup_x |k''(x)| = k''(\Theta'') < \infty$. The density itself is bounded as well: $\sup_x |k(x)| = k(\Theta) < \infty$. □

5.1.2 Consistency of the estimator

As a first result, we establish the consistency of the estimator. Clearly, like all kernel density estimators, it is not unbiased in finite samples. We consider each term of the decomposition (5.1) individually and assume Conditions 5.1.2 to hold for the next lemmas. The main term nearly gives us the desired result:

Lemma 5.1.3 (Convergence in probability of A_1). *It holds*

$$\boldsymbol{E}[A_1] = \boldsymbol{E}\left[\frac{1}{nh\varepsilon}\sum_{t=1}^{n} K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] = k(y) f(0) + \mathcal{O}\left(h^2 + \varepsilon^2\right)$$

$$\boldsymbol{Var}[A_1] = \mathcal{O}\left(\frac{1}{nh\varepsilon}\right)$$

The prefactor norms this term in the desired way:

Lemma 5.1.4 (Convergence in probability of A_2).

$$A_2 = \frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right) = f(0) + \mathcal{O}_P\left(\varepsilon^2 + \frac{1}{\sqrt{n\varepsilon}}\right)$$

The following terms that result out of Equation (5.1) vanish asymptotically:

Lemma 5.1.5 (Convergence in probability of B_1).

$$B_1 = \frac{\varphi - \hat{\varphi}}{h}\frac{1}{nh\varepsilon}\sum_{t=1}^{n} X_{t-1} K'\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) = \mathcal{O}_P\left(\frac{\varepsilon^2}{\sqrt{n}} + \frac{\varepsilon^3}{h\sqrt{n}} + \frac{\sqrt{\varepsilon}}{nh^{1.5}}\right)$$

Lemma 5.1.6 (Convergence in probability of B_2).

$$\begin{aligned} B_2 &= \left(\frac{\varphi - \hat{\varphi}}{h}\right)^2 \frac{1}{nh\varepsilon}\sum_{t=1}^{n} X_{t-1}^2 K''\left(\frac{b_t X_{t-1} + e_t - y}{h} + \vartheta\frac{\varphi - \hat{\varphi}}{h} X_{t-1}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \\ &= \mathcal{O}_P\left(\frac{\varepsilon^{1.5}}{n^{1.5} h^3}\right), \end{aligned}$$

The proof of Lemmas 5.1.3 through 5.1.6 is delayed to Section 5.6, so that we can finally state the consistency of the estimator:

Theorem 5.1.7. *Under Conditions 5.1.2 the estimator $\hat{k}(y)$, given in Definition 5.1.1, is consistent for $k(y)$ with*

$$\boldsymbol{Bias}\left[\hat{k}(y)\right] = \mathcal{O}\left(h^2 + \varepsilon^2\right)$$

$$\boldsymbol{Var}\left[\hat{k}(y)\right] = \mathcal{O}\left(\frac{1}{nh\varepsilon}\right).$$

Proof. The assertion directly follows from Lemmas 5.1.3 through 5.1.6 and the decomposition (5.1) of the estimator. □

5.1.3 Asymptotic distribution of the estimator

Having established the asymptotic consistency of the estimator, we can draw some conclusions about its asymptotic distribution.

Theorem 5.1.8 (CLT: Asymptotic Normality)**.** *Under Conditions 5.1.2, the following convergence holds true:*

$$\sqrt{nh\varepsilon}\left(\hat{k}_{n,\varepsilon}(y) - k(y)\right) \stackrel{n\to\infty}{\longrightarrow} \mathcal{N}(B,\tau^2) \quad i.D.$$

with

$$\begin{aligned}
B &= \sqrt{nh^5\varepsilon}K_1 + \sqrt{nh\varepsilon^5}K_2 + \sqrt{nh^3\varepsilon^3}K_3 + \mathcal{O}\left(\sqrt{nh^7\varepsilon} + \sqrt{nh\varepsilon^7}\right)\\
\tau^2 &= \frac{\sigma^2}{f(0)^2}\\
\sigma^2 &= k(y)f(0)\int K^2(u)\,du\int W^2(v)\,dv + \mathcal{O}\left(h^2+\varepsilon\right)\\
|K_1| &\leqslant \frac{1}{2f(0)}\int K(u)u^2\,du\int W(v)\,dvf(0)|k''(\Theta'')|\int h(b)\,db\\
|K_2| &\leqslant \frac{1}{2f(0)}\int K(u)\,du\int W(v)v^2\,dvk(y)|f''(\Gamma'')| + \frac{1}{f(0)}f''(\Gamma'')\int W(v)v^2\,dv\\
&\qquad + \frac{1}{2f(0)}\int K(u)\,du\int W(v)v^2\,dvf(0)|k''(\Theta'')|\int b^2h(b)\,db\\
|K_3| &\leqslant \int \frac{1}{f(0)}K(u)|u|\,du\int W(v)|v|\,dvf(0)|k''(\Theta')|\int |b|h(b)\,db
\end{aligned}$$

For the proof we refer to Section 5.6. □

5.1.4 Optimal choice of the bandwidth

We want to complete the considerations of this density estimator with some remarks about the optimal bandwidths h and ε. To determine an optimal bandwidth, we consider the *Mean Square Error (MSE)* that takes into account the bias and the variance of the estimator. By Theorem 5.1.7 we have

$$MSE_{h,\varepsilon}(y) = c_1\frac{1}{nh\varepsilon} + c_2h^4 + c_3\varepsilon^4 + o\left(\frac{1}{nh\varepsilon} + h^4 + \varepsilon^4\right).$$

We consider a bandwidth to be an optimal bandwidth if it minimizes the MSE. This means that both the variance and the squared bias are then of the same order of magnitude. Let us assume that $\varepsilon = h^\alpha$, for $\alpha > 0$. Then, we can conclude:

Theorem 5.1.9. *The optimal choice of the bandwidths h and ε in the sense that the mean square error is minimized is given by $h^* = \mathcal{O}\left(n^{-\frac{1}{6}}\right)$ and $\varepsilon^* = \mathcal{O}\left(n^{-\frac{1}{6}}\right)$, yielding $MSE^*_{h,\varepsilon}(y) = \mathcal{O}\left(n^{-\frac{2}{3}}\right)$.*

The proof is deferred to Section 5.6. □

These bandwidths meet the Conditions 5.1.2. Next, we turn to estimating the density of the disturbance noise.

5.2 Disturbance parameter via simple residuals

Using the ideas of the previous chapters we can define an estimator for the density by

Definition 5.2.1.

$$\hat{h}_{n,M}(y) = \frac{1}{\sum_{t=1}^{n} \mathbb{1}_{\{|X_{t-1}| \geqslant M\}}} \sum_{t=1}^{n} \frac{1}{k} G\left(\frac{\frac{X_t}{X_{t-1}} - \hat{\varphi} - y}{k}\right) \mathbb{1}_{\{|X_{t-1}| \geqslant M\}}$$

where G is a kernel and $k = k(n) \stackrel{n\to\infty}{\longrightarrow} 0$ a bandwidth and $M = M(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ the parameter indicating that a observation is large.

As in the previous case, the kernel can be splitted up into

$$G\left(\frac{\frac{X_t}{X_{t-1}} - \hat{\varphi} - y}{k}\right) = G\left(\frac{b_t - \frac{e_t}{X_{t-1}} - y}{k}\right) + \frac{\varphi - \hat{\varphi}}{k} G'\left(\frac{b_t - \frac{e_t}{X_{t-1}} - y}{k}\right) + \frac{\varphi - \hat{\varphi}}{k} G''(\xi) \tag{5.2}$$

to show the asymptotic consistency of the estimator:

Theorem 5.2.2. *For the estimator $\hat{h}_{n,M}(y)$ holds under Conditions 5.1.2 that*

$$\boldsymbol{Bias}\left[\hat{h}_{n,M}(y)\right] = \mathcal{O}\left(k^2 + \frac{1}{\sqrt{n}}\right)$$
$$\boldsymbol{Var}\left[\hat{h}_{n,M}(y)\right] = \mathcal{O}\left(\frac{1}{nk}\right)$$

Proof. Using the expansion (5.2), this result can be obtained directly by standard methods similar to Theorem 5.1.7. □

Although we have these asymptotic results for the estimator, it is equipped with some problems. Clearly, this function is not continuous in X_{t-1} what yields some problmes in determining the asymptotic distribution of the estimator. Due to the structure that all large observations are considered, we cannot introduce a second kernel like in the last case. Furthermore, typically the observations of the stationary process X_t are mostly in a certain neighborhood of zero. Hence, for large n only a small number of observations is considered, especially if the density of the stationary distribution of the process decays fast. To circumvent this fact, another approach is considered in the next section.

5.3 Disturbance parameter via deconvolution

5.3.1 Derivation of the estimator

If we want to avoid the problems of the previous section, the technique of deconvolution promises some success. The setting is as follows: If we assume that three random variables meet the relationship $u_t = B_t + e_t$, that realizations $\hat{u}_1, \ldots, \hat{u}_n$ of u_t are given, and that the density of e_t is known, we can estimate the density of B_t if e_t and B_t are independent by standard deconvolution methods.

In our case, B_t is still a composition of two random variables: $B_t = b_t X_{t-1}$, but we always know the value of X_{t-1}. Thus, we can proceed similar to the estimation procedure for the density of the innovation noise and just take these values of $\hat{u}_t$ that "belong" to a value near one of X_{t-1}. Then, $B_t = b_t X_{t-1}$ is equal to b_t multiplicatively disturbed with some noise: $B_t = b_t(1+\delta_t)$. The noise is approximately one for small δ and hence B_t is approximately equal to b_t for small δ_t.
With $\hat{u}_t^n = X_t - \hat{\varphi} X_{t-1}$ and G a kernel function, an estimator for the density g of $u_t = b_t + e_t$ is given by

$$\hat{g}(s) = \frac{1}{\sum_{j=1}^n \mathbb{1}_{\{1-\delta(n)\leqslant X_{j-1}\leqslant 1+\delta(n)\}}} \sum_{j=1}^n \frac{1}{k} G\left(\frac{s-\hat{u}_j^n}{k}\right) \mathbb{1}_{\{1-\delta(n)\leqslant X_{j-1}\leqslant 1+\delta(n)\}}.$$

Again, it is clear that this function is not continuous in X_{t-1}, hence we can again introduce a second kernel function that ensures that we consider only these values of u_t that correspond to a value of X_{t-1} near one:

$$\hat{g}(s) = \frac{1}{\frac{1}{\delta}\sum_{j=1}^n V\left(\frac{X_{j-1}-1}{\delta}\right)} \sum_{j=1}^n \frac{1}{k\delta} G\left(\frac{s-\hat{u}_j^n}{k}\right) V\left(\frac{X_{j-1}-1}{\delta}\right)$$

Further, k and δ are the bandwidths that we use for the kernels G and V, respectively, and $k, \delta \stackrel{n\to\infty}{\longrightarrow} 0$. We can extend this and consider not only an interval around one for X_{t-1}, but also an interval around $a \neq 0$, depending on the distribution of X_t. We have to choose such a number a that we have a sufficient large number of values of X_{t-1} in the interval:

$$\hat{g}(s) = \frac{1}{\frac{1}{\delta}\sum_{j=1}^n V\left(\frac{X_{j-1}-a}{\delta}\right)} \sum_{j=1}^n \frac{1}{k\delta} G\left(\frac{s-\hat{u}_j^n}{k}\right) V\left(\frac{X_{j-1}-a}{\delta}\right) \tag{5.3}$$

Taking an interval around $a \neq 0$ for the values of X_{t-1} means that we consider observations of $u_t = ab_t + e_t, \quad a \neq 0$. We therefore estimate the density $h_{ab_t}(x)$ of ab_t and have to transform this to the density $h(x)$ of b_t via the relation

$$h_{b_t}\left(\frac{x}{a}\right) = |a| h_{ab_t}(x). \tag{5.4}$$

Standard deconvolution methods require the density of e_t to be known, but we only have an estimated density for the innovation parameters. More exactly, the methods require the characteristic function of e_t to be known. However, we can simply replace the characteristic function by an estimate.
We denote by ϕ_X the characteristic function of a random variable X and recall that $(e_t)_{t\in\mathbb{Z}}$ are i.i.d. with density $k(\cdot)$ and that $(b_t)_{t\in\mathbb{Z}}$ are independent of $(e_t)_{t\in\mathbb{Z}}$ with density $h(\cdot)$. For a random variable b_t with density $h(\cdot)$ and characteristic function ϕ_b we have the relations

$$h(z) = \frac{1}{2\pi}\int \mathrm{e}^{-itz}\phi_b(t)\mathrm{dt} \quad \text{and} \quad \phi_b(t) = \int \mathrm{e}^{itx} h(x)\mathrm{dx};$$

and since we know further

$$u_t = b_t + e_t \quad \Leftrightarrow \quad \phi_u(t) = \phi_b(t)\phi_e(t)$$

if b_t and e_t are independet, we conclude for independent b_t and e_t

$$h(z) = \frac{1}{2\pi} \int \mathrm{e}^{-itz} \frac{\phi_u(t)}{\phi_e(t)} \mathrm{d}t \tag{5.5}$$

if the characteristic function $\phi_e(t) > 0$ is positive for all t.
Having a kernel density estimator $\hat{g}$ for the density g of some random variable U we can easily construct a kernel density estimator for its characteristic function with a kernel K,

$$\hat{\phi}_U(t) = \int \mathrm{e}^{ity} \hat{g}(y) \mathrm{d}y = \frac{1}{nh} \sum_{r=1}^{n} \int \mathrm{e}^{ity} K\left(\frac{\hat{u}_r^n - y}{h}\right) \mathrm{d}y,$$

so that we can define the following estimator:

Definition 5.3.1. *In an RCA model, with $\hat{u}_t^n = X_t - \hat{\varphi} X_{t-1}$, an estimator for the characteristic function of the innovation noise is given by*

$$\hat{\phi}_e(t) = \frac{1}{\frac{1}{n\varepsilon}\sum_{r=1}^{n} W\left(\frac{X_{r-1}}{\varepsilon}\right)} \frac{1}{nh\varepsilon} \sum_{r=1}^{n} \int \mathrm{e}^{ity} K\left(\frac{y - \hat{u}_r^n}{h}\right) dy\, W\left(\frac{X_{r-1}}{\varepsilon}\right)$$

where K and W are two kernel functions and the bandwidths are $h = h(n) \stackrel{n\to\infty}{\longrightarrow} 0$ and $\varepsilon = \varepsilon(n) \stackrel{n\to\infty}{\longrightarrow} 0$.

Theorem 5.3.2. *Under the Conditions 5.1.2, the estimator $\hat{\phi}_e(t)$ given in Definition 5.3.1 is consistent for $\phi_e(t)$ and it holds:*

$$\hat{\phi}_e(t) = \left(\phi_e(t) + \frac{1}{2}h^2t^2 \int u^2 K(u)\, du + \mathcal{O}_P\left(h^3t^3 + \frac{1}{\sqrt{nh\varepsilon}}\right)\right)\left(1 + \mathcal{O}_P\left(\varepsilon^2\right)\right)$$

The proof is deferred to Section 5.6. □

Using this result together with Equations (5.3) and (5.4) as well as Definition 5.1.1 we obtain from Equation (5.5) the estimator for $h\left(\frac{z}{a}\right) \equiv h_{b_t}\left(\frac{z}{a}\right)$:

Definition 5.3.3. *With the aforementioned notation and with $a \neq 0$, an estimator for the density $h\left(\frac{z}{a}\right)$ of the disturbance noise in an RCA model is given by*

$$\begin{aligned}\hat{h}\left(\frac{z}{a}\right) &= |a| \frac{1}{2\pi} \int \mathrm{e}^{-itz} \frac{\hat{\phi}_u(t)}{\hat{\phi}_e(t)} dt \\ &= \frac{1}{n\delta} \sum_{j=1}^{n} \frac{|a|}{2\pi} \int e^{it(\hat{u}_j^n - z)} \frac{\phi_G(tk) V\left(\frac{X_{j-1}-a}{\delta}\right)}{\frac{1}{n\varepsilon}\sum_{t=1}^{n} \mathrm{e}^{it\hat{u}_r^n} \phi_K(th) W\left(\frac{X_{t-1}}{\varepsilon}\right)} dt \frac{\frac{1}{n\varepsilon}\sum_{r=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)}{\frac{1}{n\delta}\sum_{t=1}^{n} V\left(\frac{X_{t-1}-a}{\delta}\right)}\end{aligned} \tag{5.6}$$

where $\hat{\varphi}$ is an estimator for φ, $\hat{u}_t^n = X_t - \hat{\varphi} X_{t-1}$, and K, G, V, W are kernel functions and h, k, ε, δ are bandwidths in dependence on n. □

For further analysis we set

$$\begin{aligned}
\tilde{h}\left(\frac{z}{a}\right) &= \frac{1}{n\delta}\sum_{j=1}^{n}\frac{|a|}{2\pi}\int e^{it(b_jX_{j-1}+e_j-z)}\frac{\phi_G(tk)}{\phi_e(t)}V\left(\frac{X_{j-1}-a}{\delta}\right)\mathrm{dt}\\
A_2 &= \frac{1}{n\varepsilon}\sum_{t=1}^{n}W\left(\frac{X_{t-1}}{\varepsilon}\right)\\
A_3(a) &= \frac{1}{n\delta}\sum_{t=1}^{n}V\left(\frac{X_{t-1}-a}{\delta}\right)\\
T(z) &= \frac{1}{A_3(a)}\frac{|a|}{2\pi}\int\int \mathrm{e}^{it(y-z)}\frac{T_1(y)+T_2(y)}{\phi_e(t)}\mathrm{dydt}\\
T_1(y) &= \frac{\varphi-\hat{\varphi}}{k}\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}G'\left(\frac{b_jX_{j-1}+e_j-y}{k}\right)V\left(\frac{X_{j-1}-a}{\delta}\right)\\
T_2(y) &= \left(\frac{\varphi-\hat{\varphi}}{k}\right)^2\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}^2G''\left(\frac{b_jX_{j-1}+e_j-y}{k}+\vartheta\frac{\varphi-\hat{\varphi}}{k}X_{j-1}\right)V\left(\frac{X_{j-1}-a}{\delta}\right)\\
S(z) &= \frac{|a|}{2\pi}\int\frac{1}{n\delta}\sum_{r=1}^{n}\frac{\left(\phi_G(tk)+\frac{\varphi-\hat{\varphi}}{k}X_{r-1}F_{G'}(tk)+\left(\frac{\varphi-\hat{\varphi}}{k}X_{r-1}\right)^2\mathrm{e}^{\vartheta\frac{\varphi-\hat{\varphi}}{k}X_{r-1}}F_{G''}(tk)\right)}{\phi_e\left(t\right)}\\
&\qquad\qquad \mathrm{e}^{it(-z+b_rX_{r-1}+e_r)}\frac{\hat{\phi}_e\left(t\right)-\phi_e\left(t\right)}{\hat{\phi}_e\left(t\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\\
S_1(z) &= \frac{|a|}{2\pi}\int\frac{1}{nk\delta}\sum_{r=1}^{n}\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\\
S_2(z) &= S(z)-S_1(z)\\
J_2 &= \frac{1}{16\pi}\int u^2K(u)\mathrm{du}\int\int\mathrm{e}^{-isu}s^2\phi_G(s)\mathrm{dsdu}.
\end{aligned}$$

We will refer to these terms frequently in this section. We note that

$$\int\mathrm{e}^{ity}G\left(\frac{b_rX_{r-1}+e_r-y}{h}\right)\mathrm{dy}=k\int\mathrm{e}^{it(b_rX_{r-1}+e_r-vk)}G(v)\mathrm{dv}=k\int\mathrm{e}^{it(b_rX_{r-1}+e_r)}\phi_G(tk)$$

and obtain by a taylor expansion for the kernel as in Equation (5.1) from Equation (5.6)

$$\begin{aligned}
\hat{h}\left(\frac{z}{a}\right) &= \frac{1}{A_3(a)}|a|\frac{1}{2\pi}\int\frac{1}{nk}\sum_{r=1}^{n}\int\mathrm{e}^{it(y-z)}G\left(\frac{\hat{u}_r-y}{h}\right)\mathrm{dy}\frac{1}{\phi_e(t)}\mathrm{dt}V\left(\frac{X_{j-1}-a}{\delta}\right)\\
&\quad+\frac{1}{A_3(a)}|a|\frac{1}{2\pi}\int\frac{1}{nk}t^3\sum_{r=1}^{n}\frac{\int\mathrm{e}^{it(y-z)}G\left(\frac{\hat{u}_r-y}{h}\right)\mathrm{dy}V\left(\frac{X_{j-1}-a}{\delta}\right)}{\phi_e(t)\,\hat{\phi}_e(t)}\mathrm{dt}\\
&= \frac{1}{A_3(a)}|a|\frac{1}{2\pi}\int\frac{1}{nk}\sum_{r=1}^{n}\int\mathrm{e}^{it(y-z)}G\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)\mathrm{dy}\frac{V\left(\frac{X_{j-1}-a}{\delta}\right)}{\phi_e(t)}\mathrm{dt}+T(z)+S(z)\\
&= \frac{1}{A_3(a)}\tilde{h}\left(\frac{z}{a}\right)+T(z)+\frac{S(z)}{A_3(a)} \qquad (5.7)
\end{aligned}$$

and will see that the key term is $\tilde{h}\left(\frac{z}{a}\right)$.
To make some conclusions about the consistency or the asymptotic distribution of the estimator we have to put some further assumptions, especially on the distribution of the innovation noise. We proceed like Fan (1991) and separate two cases.

1. The smooth case: this means, the characteristic function of the innovation noise behaves like

$$\begin{aligned} \phi_e(u) &\overset{u\to\infty}{\longrightarrow} \tilde{c}|u^{-\beta}| \\ \text{i.e.} \quad \frac{1}{\phi_e(u)} &\leqslant c|u|^{\beta} \quad \forall |u| \geqslant M \end{aligned}$$

   for constants $\tilde{c}, c, \beta \geqslant 0$ and a sufficient large but fixed M. In this case we have to choose the kernel function G such that

$$\int |\phi_G(t)|\ |t|^{\beta}\mathrm{dt} < \infty.$$

2. The supersmooth case: this means, the characteristic function of the innovation noise behaves like

$$\begin{aligned} c_2 \geqslant \phi_e(u)|u^{-\beta_0}|\mathrm{e}^{|u|^{\beta}/\gamma} &\geqslant c_1 \quad \forall |u| \geqslant M \\ \text{i.e.} \quad \frac{1}{c_2}|u|^{-\alpha}\mathrm{e}^{|u|^{\beta}/\gamma} \leqslant \frac{1}{\phi_e(u)} &\leqslant \frac{1}{c_1}|u|^{-\alpha}\mathrm{e}^{|u|^{\beta}/\gamma} \quad \forall |u| \geqslant M \end{aligned} \tag{5.8}$$

   for constants $c_1, c_2, \beta, \gamma \geqslant 0$, $\alpha \in \mathbb{R}$, and a sufficient large but fixed M. In this case, we have to choose the kernel function G such that its characteristic function vanishes outside a compact interval, for example outside of $[-1, 1]$.

First, we consider just the smooth case.

5.3.2 The smooth case

We assume the following conditions to hold for the remainder of this chapter:

Conditions 5.3.4.

- Conditions 5.1.2 hold.
- We consider two non-negative symmetric and centered kernel functions G and V that are two times continuously differentiable. V as well as the second derivative of G is assumed to be bounded: $\sup_x |G''(x)| = G''(\Lambda'')$. V is Lipschitz continuous.
- For a function G we denote its inverse Fouriertransform by F_G
- The bandwidths and kernel functions further satisfy $k = k(n) \overset{n\to\infty}{\longrightarrow} 0$, $\delta = \delta(n) \overset{n\to\infty}{\longrightarrow} 0$, $nk^{2\beta+1}\delta \overset{n\to\infty}{\longrightarrow} \infty$, $nk^{\beta+1.5}\delta \overset{n\to\infty}{\longrightarrow} \infty$, $k^{\beta}\sqrt{nk^5\delta} + k^{\beta}\sqrt{nk\delta^5} \overset{n\to\infty}{\longrightarrow} c < \infty$, $\frac{h^4}{nk^{4\beta+5}\delta} \overset{n\to\infty}{\longrightarrow} 0$, $\frac{h^2}{k^2} \overset{n\to\infty}{\longrightarrow} 0$, $\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}\left(\frac{1}{n} + k^2 + \delta^2 + \varepsilon^2 + \frac{h^3}{k^3}\right) \overset{n\to\infty}{\longrightarrow} c < \infty$, $\phi_e(t) \overset{|t|\to\infty}{\longrightarrow} c|t^{\beta}|$, $c > 0$, $\int G^i(u)|u^j|\mathrm{du} < \infty$, $(i = 1, 2,\ j = 0, 1, 2$ and $j = 0$ for $i = 3, 4)$, $\int |G'(u)|\ |u^j|\mathrm{du} < \infty$, $(j = 0, 1, 2)$, $\int V^i(u)\mathrm{du} < \infty$, $(i = 3, 4)$, $\int V^i(u)|u^j|\mathrm{du} < \infty$, $(i = 1, 2,\ j = 0, 1, 2, 3, 4, 5)$, $\int |\phi_G(t)||t^{2\beta+3}|\mathrm{dt} < \infty$, $\int |F_{G'}(t)||t^{2\beta+3}|\mathrm{dt} < \infty$, $\int |F_{G''}(t)||t^{2\beta+3}|\mathrm{dt} < \infty$, $\int |\phi_G^4(t)||t^{4\beta-1}|\mathrm{dt} < \infty$.

- The density of f of X is continuous in a neighborhood of 0 and of $a \neq 0$, further, $f(0) > 0$ and $f(a) > 0$.
- The density of b_t exists, is denoted by $h(\cdot)$ and is two times continuous differentiable with bounded first and second derivatives: $\sup_x |h''(x)| = h''(\Psi'') < \infty$. The density itself is bounded as well: $\sup_x |h(x)| = h(\Psi) < \infty$.
- For ease of calculation we assume further without loss of generalization $\beta \geqslant 1$. $\square$

One kernel G that satisfies these conditions is, for example, the normal distribution while the white noise $(e_t)_{t\in\mathbb{Z}}$ could follow, for example, the double-exponential, the gamma, or the triangular distribution (see also Fan (1991)), however, we assumed $(e_t)_{t\in\mathbb{Z}}$ to be centered. The double-exponential distribution, for example, has parameter $\beta = 2$.

The Conditions 5.3.4 ensure that, for M large enough,

$$\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \leqslant \begin{cases} \frac{\sup_{t\in[-Mk,Mk]} \phi_G(t)}{\inf_{t\in[-Mk,Mk]} \phi_e\left(\frac{t}{k}\right)} & \text{if } t \in [-Mk, Mk] \\ c\frac{|t|^\beta}{k^\beta}|\phi_G(t)| & \text{else,} \end{cases} \quad \text{and thus} \quad \int \left|\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\right| \mathrm{d}t < \infty.$$

Consistency of the estimator

We first state two Lemmas to establish the consistency.

Lemma 5.3.5. *Under Conditions 5.3.4 it holds for the main term of the decomposition (5.7):*

$$\begin{aligned} \boldsymbol{E}\left[\tilde{h}\left(\frac{z}{a}\right)\right] &= h\left(\frac{z}{a}\right) f(a) + \mathcal{O}\left(k^2 + \delta^2\right) \\ \boldsymbol{Var}\left[\tilde{h}\left(\frac{z}{a}\right)\right] &= \mathcal{O}\left(\frac{1}{nk^{2\beta+1}\delta}\right) \end{aligned}$$

For the proof we refer to Section 5.6. $\square$
One of the norming factors is evaluated in Lemma 5.1.4, for the other one we have:

Lemma 5.3.6 (Convergence of A_3)**.**

$$\boldsymbol{E}[A_3(a)] = \boldsymbol{E}\left[\frac{1}{n\delta}\sum_{j=1}^{n} V\left(\frac{X_{j-1} - a}{\delta}\right)\right] = f(a) + \mathcal{O}(\delta^2)$$

$$\boldsymbol{Var}[A_3(a)] = \mathcal{O}\left(\frac{1}{n\delta}\right)$$

Proof. The same computations as in Lemma 5.1.4. $\square$

Lemma 5.3.7 (Convergence of $T(z)$)**.**

$$T(z) = \frac{1}{A_3}\frac{|a|}{2\pi}\int\int \mathrm{e}^{it(y-z)}\frac{T_1(y) + T_2(y)}{\phi_e(t)}dydt = \mathcal{O}_P\left(\frac{1}{k\sqrt{n}} + \frac{1}{nk^{\beta+1}\sqrt{k\delta}}\right)$$

where $T_1(y)$ and $T_2(y)$ are given after Definition 5.3.3.

The proof is deferred to Section 5.6. ☐

Lemma 5.3.8. *For the term $S(z) = S_1(z) + S_2(z)$ it holds that:*

$$\boldsymbol{E}[S(z)] = \frac{h^2}{k^2}J_2 + \mathcal{O}\left(\frac{1}{n} + k^2 + \delta^2 + \varepsilon^2 + \frac{h^3}{k^3}\right)\left(1 + \mathcal{O}\left(\frac{h^2}{k^2} + \varepsilon^2 + \frac{1}{\sqrt{nh\varepsilon}}\right)\right)$$

$$\text{and} \quad \boldsymbol{E}\left[S_1^2(z)\right] = \mathcal{O}\left(\frac{h^4}{nk^{4\beta+5}\delta}\right), \quad \boldsymbol{E}\left[S_2^2(z)\right] = \mathcal{O}\left(\frac{h^4}{n^2k^{4\beta+6}\delta}\right).$$

The proof is delayed to Section 5.6. ☐
With these preparations we can state the main result:

Theorem 5.3.9. *If Conditions 5.1.2 and 5.3.4 are fulfilled, the estimator $\hat{h}\left(\frac{z}{a}\right)$ given in Definition 5.3.3 is consistent for $h\left(\frac{z}{a}\right)$ with*

$$\hat{h}\left(\frac{z}{a}\right) = h\left(\frac{z}{a}\right) + \mathcal{O}_P\left(k^2 + \delta^2 + \varepsilon^2 + \frac{h^2}{k^2} + \frac{1}{k^\beta\sqrt{nk\delta}} + \frac{h^2}{k^{2\beta+2}\sqrt{nk\delta}}\right)$$

Proof. Lemmas 5.1.4 and 5.3.5 through 5.3.8 directly yield the assertion with Slutsky via the decomposition (5.7). ☐

Asymptotic distribution of the estimator

Having established the asymptotic consistency of the estimator, we can draw some conclusions about its asymptotic distribution.

Theorem 5.3.10. *Under Conditions 5.3.4 the following convergence holds true:*

$$k^\beta\sqrt{nk\delta}\left(\frac{\tilde{h}_n\left(\frac{z}{a}\right)}{A_3(a)} - h\left(\frac{z}{a}\right)\right) \xrightarrow{n\to\infty} \mathcal{N}(\tilde{B}, \tilde{\tau}^2) \quad i.D.$$

with

$$\begin{aligned}
\tilde{B} &= k^\beta\sqrt{nk\delta}\left(k^2K_1 + k\delta K_2 + \delta^2K_3 + \mathcal{O}\left(k^3 + \delta^3\right)\right) = k^\beta\sqrt{nk\delta}B_1 \\
\tilde{\tau}^2 &= \frac{\tilde{\sigma}^2}{f(a)^2} \\
\tilde{\sigma}^2 &= f(a)a^2\int V^2(v)\,dv\int k(z+ba)h(b)\,db\int t^{2\beta}\phi_G^2(t)\,dt + \mathcal{O}\left(k^2+\delta\right) \\
|K_1| &\leqslant \frac{1}{2}|a|h''(\Psi'')\int G(u)u^2\,du \\
|K_2| &\leqslant \int |u|G(u)\,du\int |v|V(v)\,dv\left(\left|\frac{z}{a}\right|\frac{h''(\Psi'')}{f(a)} + \left|h'\left(\frac{z}{a}\right)\right|\left(\left|\frac{f'(a)}{af(a)}\right| + \frac{1}{|a|^3}\right)\right) \\
|K_3| &\leqslant \int V(v)v^2\,dv\left(\frac{|z|}{a^4}|h'\left(\frac{z}{a}\right)| + \frac{z^2}{2a^2}|h''(\Psi'')| + \frac{|z|}{a^2}\frac{|f'(a)|}{f(a)}|h'\left(\frac{z}{a}\right)| + \frac{f''(\Gamma'')}{2f(a)}h\left(\frac{z}{a}\right)\right)
\end{aligned}$$

Having established the consistency of the estimator, the proof is basically the same as for Theorem 5.1.8 and therefore ommited. ☐

Lemma 5.3.11. *Under Conditions 5.3.4 the following convergence holds true:*

$$\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}\left(S(z)-\frac{h^2}{k^2}J_2\right)\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(B_S,\sigma_S^2\right)\quad i.D.$$

with

$$\begin{aligned}B_S &= \frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}\cdot\mathcal{O}\left(\frac{1}{n}+k^2+\delta^2+\varepsilon^2+\frac{h^3}{k^3}\right)=\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}B_2\\ \sigma_S^2 &= \frac{a^2}{4\pi^2}\int s^{4+4\beta}\phi_G^2(s)\,ds f(a)\int V^2(v)\,dv\int k(z-ba)h(b)\,db+o(1).\end{aligned}$$

For the proof we refer to Section 5.6. □

Lemma 5.3.12. *Let Conditions 5.3.4 hold and let $\frac{k^{\beta+2}}{h^2}\neq c\quad\forall\, c\in\mathbb{R}$. Then, the following convergence holds true:*

$$\min\left\{\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2},k^\beta\sqrt{nk\delta}\right\}\left(\hat{h}_n\left(\frac{z}{a}\right)-h\left(\frac{z}{a}\right)-\frac{h^2}{k^{\beta+2}}\frac{J_2}{f(a)}\right)\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(B_h,\tau_h^2\right)\quad i.D.$$

with

$$\begin{aligned}B_h &= \min\left\{\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2},k^\beta\sqrt{nk\delta}\right\}(B_1+B_2)\\ \tau_h^2 &= \mathbb{1}_{\left\{\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}<k^\beta\sqrt{nk\delta}\right\}}\tau_s^2+\mathbb{1}_{\left\{\frac{k^{2\beta+2}\sqrt{nk\delta}}{h^2}>k^\beta\sqrt{nk\delta}\right\}}\tilde{\tau}^2\end{aligned}$$

Proof. We recall the decomposition

$$\hat{h}_n\left(\frac{z}{a}\right)-h\left(\frac{z}{a}\right)-\frac{h^2}{k^2}\frac{J_2}{f(a)}=\frac{\tilde{h}_n\left(\frac{z}{a}\right)}{A_3}-h\left(\frac{z}{a}\right)+T(z)+\frac{S(z)}{A_3}-\frac{h^2}{k^2}\frac{J_2}{f(a)}$$

from Equation (5.7) and the assertion directly follows from Lemmas 5.3.6 and 5.3.7 together with Theorem 5.3.10 and Lemma 5.3.11. □

With these results we can finally state the main result, the asymptotic normality of $\hat{h}(\cdot)$.

Theorem 5.3.13. *Let Conditions 5.3.4 hold and let for an arbitrary $C\in(0,\infty)$: $\frac{h^2}{k^{2+\beta}}\overset{n\to\infty}{\longrightarrow}C$. Then, the following central limit theorem holds true for the density estimator $\hat{h}(\cdot)$:*

$$k^\beta\sqrt{nk\delta}\left(\hat{h}_n\left(\frac{z}{a}\right)-h\left(\frac{z}{a}\right)-\frac{h^2}{k^2}\frac{J_2}{f(a)}\right)\overset{n\to\infty}{\longrightarrow}\mathcal{N}\left(B_M,\sigma_M^2\right)\quad i.D.$$

with

$$\begin{aligned}B_M &= k^\beta\sqrt{nk\delta}\left(B_1+C\frac{B_2}{f(a)}\right)\\ \sigma_M^2 &= \frac{a^2}{f(a)}\int V^2(v)\,dv\int k(z+ba)h(b)\,db\int t^{2\beta}\phi_G^2(t)\,dt\\ &\quad+C^2\frac{a^2}{4\pi^2}\frac{1}{f(a)^2}\int s^{4+4\beta}\phi_G^2(s)\,ds\int V^2(v)\,dv\int k(z-ba)h(b)f(x)\,db\\ &\quad+\frac{2C(-1)^\beta}{f(a)^2}\frac{|a|}{4\pi^2}h\left(\frac{z}{a}\right)f(a)\int\int\phi_e\left(\frac{w}{k}\right)\mathrm{e}^{iwu}\,du\,dw\int V^2(v)\,dv\int s^{2\beta+2}\phi_G^2(s)\,ds+o(1).\end{aligned}$$

The proof is given in Section 5.6. □

Remark 5.3.14. *Asymptotically, it is only necessary that the density f is positive in zero and in a and that it is continuous in a neighborhood of these two points. For practical use, however, the parameter a should be chosen carefully such that there is high chance to have a sufficient large number of observations around a. In addition, it is important that a is sufficiently far away from zero, that means that it is not too small in absolute value for the estimate to be robust.*
To minimize the asymptotic variance, it is necessary to choose a as small as possible. The minimal variance is obtained for $a = 0$ which is not allowed and it should also be not too small for the estimate to be robust. The different bias terms are minimized for values of a in $[1, 2]$. Hence, without further analyzing the terms a good choice of a might be $a = 1$. □

Optimal choice of the bandwidth

To complete these considerations, we want to state some results on the choice of the bandwidths. Again, we consider a bandwidth to be an optimal bandwidth if it minimizes the *MSE*. Let us assume the following relation: $\delta = k^\alpha$, for $\alpha > 0$. Then we obtain:

Theorem 5.3.15. *The optimal choice of the bandwidths k and δ in the sense that the mean square error between $\tilde{h}(\cdot)$ and $h(\cdot)$ is minimized is given by $k^* = \mathcal{O}\left(n^{-\frac{1}{2\beta+6}}\right)$ and $\delta^* = \mathcal{O}\left(n^{-\frac{1}{2\beta+6}}\right)$ yielding $MSE^*_{k,\delta}(y) = \mathcal{O}\left(n^{-\frac{2}{\beta+3}}\right)$.*
However, minimizing the MSE between $\hat{h}(\cdot)$ and $h(\cdot)$ gives a slightly larger bandwidth for large β: In this case k and h have to be choosen to be $k = \mathcal{O}\left(h^{\frac{2}{\beta+2}}\right)$ and the optimal bandwidths are $k^* = \mathcal{O}\left(n^{-\frac{1}{2\beta+6}}\right)$ and $\delta^* = \mathcal{O}\left(n^{-\frac{1}{2\beta+6}}\right)$ yielding $MSE^*(y) = \mathcal{O}\left(n^{-\frac{2}{\beta+3}}\right)$ for $\beta \leqslant \frac{4}{3}$, and $k^* = \mathcal{O}\left(n^{-\frac{1}{5\beta+2}}\right)$ and $\delta^* = \mathcal{O}\left(n^{-\frac{1}{5\beta+2}}\right)$ yielding $MSE^*(y) = \mathcal{O}\left(n^{-\frac{3\beta}{5\beta+2}}\right)$ for $\beta > \frac{4}{3}$.

The proof is deferred to Section 5.6. □

Remark 5.3.16. *Depending on the parameter β one has to check if the bandwidths given by Theorem 5.3.15 meet the Conditions 5.3.4. For $\beta = 2$ the optimal bandwidth k^* yields the optimal bandwidth h^* determined in Theorem 5.1.9. Otherwise, one has to decide which error term should be minimized.* □

5.3.3 The supersmooth case

The assumptions that we put on the characteristic function of the innovation noise (see Equation (5.8)) yield that in the supersmooth case it has to behave like

$$\frac{1}{\phi_e\left(\frac{t}{k}\right)} \sim c\frac{k^\alpha}{|t|^\alpha}\mathrm{e}^{\frac{t^\beta}{k^\beta\gamma}} \qquad \text{for } \frac{|t|}{k} \to \infty$$

with $\gamma > 0$, $\beta > 0$, and $\alpha \in \mathbb{R}$.

Under some regularity conditions on the kernels and bandwidths, the following results hold true. One kernel G that satisfies these conditions is, for example, the N.N. distribution while the white noise $(e_t)_{t\in\mathbb{Z}}$ could follow, for example, the normal or the chauchy distribution or one of their mixtures (see also Fan (1991)). The normal distribution, for example, has the parameters $\beta = 2$ and $\alpha = 0$.

Theorem 5.3.17. *In the supersmooth case, the estimator* $\hat{h}\left(\frac{z}{a}\right)$ *is consistent for* $h\left(\frac{z}{a}\right)$ *with*

$$\textbf{\textit{Bias}}\left[\hat{h}\left(\frac{z}{a}\right)\right] = \mathcal{O}\left(k^2+\delta^2\right)$$
$$\textbf{\textit{Var}}\left[\hat{h}\left(\frac{z}{a}\right)\right] = \mathcal{O}\left(\frac{(\ln n)^{\frac{2}{\beta}}}{n^{1-2\vartheta}\delta}\right)$$

for the bandwidth $k = (\vartheta\gamma\ln n)^{-1/\beta}$ *and some other regularity conditions, comparable to Conditions 5.3.4.*

Proof. The proof is analogous to the computations in the smooth case and the argumentation of Fan (1991). First,

$$\mathbf{E}\left[\hat{h}\left(\frac{z}{a}\right)\right] = h\left(\frac{z}{a}\right)f(a) + \mathcal{O}\left(k^2+\delta^2\right).$$

Further,

$$\begin{aligned}
&\frac{1}{n\delta^2}\frac{a^2}{4\pi^2}\mathbf{Var}\left[\int e^{it(b_jX_{j-1}+e_j-z)}\frac{\phi_G(tk)}{\phi_e(t)}\mathrm{d}tV\left(\frac{X_{j-1}-a}{\delta}\right)\right]\\
&\leqslant \frac{1}{nk^2\delta^2}\frac{a^2}{4\pi^2}\mathbf{E}\left[\left(\int e^{-it\frac{-b_jX_{j-1}-e_j+z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_{j-1}-a}{\delta}\right)\right)^2\right]\\
&= \frac{1}{nk\delta^2}\frac{a^2}{4\pi^2}\int\int\int\left(\int e^{-itu}\frac{\phi_G(t)}{\phi_e(t)}\mathrm{d}t\right)^2\mathrm{d}uV^2\left(\frac{x-a}{\delta}\right)k(z-bx)h(b)f(x)\mathrm{dbdx}\\
&= \frac{1}{nk\delta^2}\frac{a^2}{4\pi^2}\int\int\int\frac{\phi_G^2(t)}{\phi_e^2(t)}\mathrm{d}tV^2\left(\frac{x-a}{\delta}\right)k(z-bx)h(b)f(x)\mathrm{dbdx}\\
&= \frac{a^2}{nk\delta^2}\int\int\left(c^2\left(\int_{-1}^{-MK}+\int_{Mk}^{1}\right)\frac{k^{2\alpha}}{t^{2\alpha}}\phi_G^2(t)\left(\mathrm{e}^{\frac{1}{k^\beta\gamma}}\right)^{2t^\beta}\mathrm{d}t+\mathcal{O}(k)\right)\\
&\qquad\qquad V^2\left(\frac{x-a}{\delta}\right)k(z-bx)h(b)f(x)\mathrm{dbdx}\\
&\leqslant \frac{a^2}{nk\delta}\int\int c^2\sup_{t\in[-1,1]}\phi_G^2(t)\mathrm{e}^{\frac{2}{k^\beta\gamma}}\left(\int_{-1}^{-MK}+\int_{Mk}^{1}\right)\frac{k^{2\alpha}}{t^{2\alpha}}\mathrm{d}t\\
&\qquad\qquad f(a)V^2(v)\mathrm{d}v\int k(z+ba)h(b)\mathrm{d}b+\mathcal{O}(k+\delta)\\
&= \mathcal{O}\left(\frac{\mathrm{e}^{\frac{2}{k^\beta\gamma}}}{nk^2\delta}\right),\quad \text{for } \alpha\neq\frac{1}{2},\ \text{for } \alpha=\frac{1}{2} \text{ we even obtain } \mathcal{O}\left(\frac{\mathrm{e}^{\frac{2}{k^\beta\gamma}}}{nk\delta}\right),\\
&= \mathcal{O}\left(\frac{(\ln n)^{\frac{2}{\beta}}}{n^{1-2\vartheta}\delta}\right),
\end{aligned}$$

and the covariance terms are asymptotically of smaller order than the variance terms, what can be shown similarly to Lemma 5.3.5. □

As one can see, one has to choose the parameters very carefully for the estimator to be consistent. We have to choose a logarithmic bandwidth for k because of the term $\mathrm{e}^{\frac{2}{k^{\beta}\gamma}}$ that appears in the variance of the estimator. Depending on the parameter ϑ of the bandwidth k, the bandwidth δ has to be chosen: for $\vartheta = \frac{1}{2}$, $\delta = \mathcal{O}\left((\ln n)^{-\frac{1}{\xi}}\right)$ with $0 < \xi < \beta$; for $\vartheta < \frac{1}{2}$, it can be chosen arbitrarily; $\vartheta > \frac{1}{2}$ is not allowed for the estimator to be consistent. Similar to the smooth case the asymptotic normality of the estimator can be established for the super smooth case.

5.4 Variations and practical implications

In this section, we would like to present some variations of the density estimators, that are especially important for practical implications.

5.4.1 Multipoint estimator

Instead of considering just an interval around one point a for X_{t-1} to be in, that means considering the composition $ab_t + e_t$ to estimate the density of b_t, we could choose several values $a_1, \ldots, a_N$, estimate the density $\hat{h}_i(x)$ for each of them and combine them to an estimator for $\hat{h}(x)$:

Definition 5.4.1. *For a given set $a_1, \ldots, a_N$ with $a_i \neq 0$, $|a_i|$ large enough, and $a_i \neq a_j$ $\forall i \neq j$, we set $z_i = xa_i$ and determine $\hat{h}\left(\frac{z_i}{a_i}\right)$ $\forall\, i = 1, \ldots, N$. Then, the estimator $\hat{h}^{+}(x)$ for $h(x)$ is defined by*

$$\hat{h}^{+}(x) = \frac{1}{N}\sum_{i=1}^{N} \hat{h}\left(\frac{z_i}{a_i}\right)$$

with $\hat{h}(\cdot)$ given by Definition 5.3.3. □

This estimator is especially powerful if we let vary N with n, i.e. $N \overset{n\to\infty}{\longrightarrow} \infty$, and require that the kernel V has bounded support, but one should keep in mind as well that the variance of $\hat{h}(x)$ is proportional to $\frac{a^2}{f(a)}$.

Remark 5.4.2. *The number and position of the points used for the density estimation can, for example, be determined as follows. Choose an interval $[-N^*, +N^*]$ with $f(x)$ sufficiently large for all $x \in [-N^*, +N^*]$. Set $N = \frac{N^*}{\delta}$ and choose $a_1, \ldots, a_N$ each $\neq 0$ equidistant in the whole interval $[-N^*, +N^*]$. Then, $|a_i - a_j| \geqslant \delta$. Further, choose a kernel V that is only larger than zero for $x \in (-1, +1)$. A refinement of this is the following: Choose $P \in \mathbb{N}$ as a minimum number of summands for $h^*(x)$ and set $Q = \frac{N^*}{\delta}$. Set further $N = \sqrt[n]{P^n + Q^n}$. Choose $a_1, \ldots, a_N$, each not zero, equidistant in the whole interval $[-N^*, +N^*]$. Then, $|a_i - a_j| \overset{n\to\infty}{\longrightarrow} D \geqslant \delta$ since $N \overset{n\to\infty}{\longrightarrow} Q = \frac{N^*}{\delta}$.*
With these parameters $V\left(v + \frac{a_i - a_j}{\delta}\right) > 0$ for $v \in \left(-1 - \frac{a_i - a_j}{\delta}, 1 - \frac{a_i - a_j}{\delta}\right)$, and thus, $V(v)V\left(v + \frac{a_i - a_j}{\delta}\right) > 0$ only if $v \in [-1, +1]$ and $|a_i - a_j| < 2\delta$. □

Lemma 5.4.3. *Under Conditions 5.3.4 and the points for the estimator $\hat{h}^+(x)$, given in Definition 5.4.1, determined by the rule of Remark 5.4.2 is consistent in the smooth case with*

$$\boldsymbol{Bias}[\hat{h}^+(x)] = \mathcal{O}\left(k^2 + \delta^2\right)$$
$$\boldsymbol{Var}[\hat{h}^+(x)] = \mathcal{O}\left(\frac{1}{Nnk^{2\beta+1}\delta}\right)$$

Proof. By Theorem 5.3.9 we obtain

$$\mathbf{E}\left[\hat{h}^+(x)\right] = \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N}\tilde{h}\left(\frac{z_i}{a_i}\right)\right] = h\left(x\right) + \mathcal{O}\left(k^2 + \delta^2\right),$$

and by Lemma 5.6.4 and the argumentation of the proof of Lemma 5.3.5

$$\begin{aligned}
&\mathbf{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\tilde{h}\left(\frac{z_i}{a_i}\right)\right] \\
&\quad\leqslant \frac{1}{nk^2\delta^2}\frac{1}{4\pi^2}\mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}|a_i|\int e^{it\frac{b_jX_{j-1}+e_j-z_i}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_{j-1}-a_i}{\delta}\right)\right)^2\right] \\
&\quad + \frac{2}{nk^2\delta^2}\sum_{j=1}^{n}\frac{1}{4\pi^2}\mathbf{Cov}\left[\frac{1}{N}\sum_{i=1}^{N}|a_i|\int e^{it\frac{b_1X_0+e_1-z_i}{k}}\frac{\phi_G(k)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_0-a_i}{\delta}\right),\right. \\
&\quad\qquad\left.\frac{1}{N}\sum_{i=1}^{N}|a_i|\int e^{it\frac{b_jX_{j-1}+e_j-z_i}{k}}\frac{\phi_G(k)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_{j-1}-a_i}{\delta}\right)\right] \\
&\quad = \mathcal{O}\left(\frac{1}{Nnk^{2\beta+1}\delta} + \frac{1}{nk^3}\right).
\end{aligned}$$

□

Analogous to the standard estimator, some conclusions about the asymptotic normality of this estimator can be drawn and the results can be transfered to the supersmooth case.

5.4.2 Non-negative estimator

For practical implementations, we propose two modifications of the aforementioned estimators. For ease of notation, we use the standard estimator, but the multipoint estimator could be used as well. The estimator $\hat{h}(\cdot)$ can be negative and typically will be negative in samples of small size. It is well known that the true density is non-negative, so we present two ways out.

Definition 5.4.4. *For the density estimator $\hat{h}(\cdot)$ of $h(\cdot)$ given in Definition 5.3.3, the modified density estimator $\hat{h}^{\#}(\cdot)$ is defined by*

$$\hat{h}^{\#}\left(\frac{z}{a}\right) = \hat{h}\left(\frac{z}{a}\right)\mathbb{1}_{\{\hat{h}(\frac{z}{a})\geqslant 0\}}. \tag{5.9}$$

Lemma 5.4.5. *The estimator $\hat{h}^{\#}(\cdot)$ is consistent for $h(\cdot)$ and the same conclusions about the asymptotic behavior hold as for $\hat{h}(\cdot)$.*

Proof. For n converging to infinity, the density $h(x)$ is either equal to zero (then, $\hat{h}(x)$ will converge to zero in probability) or it is greater than zero and then $\hat{h}(x)$ will converge to that value so that the indicator is positive. For all x where $h(x)$ is greater than zero, the indicator will be one for n converging to infinity. □

We note that this estimator does not necessarily integrate to one in finite samples and typically will not do so in small sample sizes. Hence, for practical implementation, the density should be normed after it is estimated.

5.4.3 Truncated estimator

Another possibility is to assume that the true density has values greater than zero only on a connected support. Then, it is most likely that the true density $h(x)$ is very small when the estimated density $\hat{h}(x)$ is negative and, when going from the center to the tail, is not growing anymore. So, our procedure is as follows: Going from the center to the tails search for these two points where the estimator is negative for the first time. Set the estimated density to zero out of these two points.

Definition 5.4.6. *Let μ be the mean of the probability distribution given by $\hat{h}(\cdot)$ and*

$$\eta_1 = \max\left\{x : \hat{h}(x) < 0, x < \mu\right\}, \quad \eta_2 = \min\left\{x : \hat{h}(x) < 0, x > \mu\right\}. \tag{5.10}$$

If necessary, we define η_1 or η_2 to be minus or plus infinity, respectively. For the density estimator $\hat{h}(\cdot)$ given in Definition 5.3.3, the modified density estimator $\hat{h}^*(\cdot)$ is defined by

$$\hat{h}^*\left(\frac{z}{a}\right) = \hat{h}\left(\frac{z}{a}\right)\mathbb{1}_{\left\{\eta_1 < \frac{z}{a} < \eta_2\right\}}. \tag{5.11}$$

Lemma 5.4.7. *The estimator $\hat{h}^*(\cdot)$ is consistent for $h(\cdot)$ and the same conclusions about the asymptotic behavior hold as for $\hat{h}(\cdot)$.*

Proof. For n converging to infinity, the density $h(x)$ is either equal to zero (then, $\hat{h}(x)$ will converge to zero in probability) or it is greater than zero and then $\hat{h}(x)$ will converge to this value so that the indicator set will be one on the whole support of $h(x)$. □

Like the previous estimator, this one will also not necessarily integrate to one in finite samples and needs to be normed as well. This last estimator can directly be used to estimate moments of the random variables. The other estimators we proposed here can be used to determine moments as well, but they need to be restricted to a compact interval like proposed in Definiton 4.3.3. For more information on density estimators that can be used to determine moments of the random variables we refer to Lemma 4.3.4.

5.5 A simulation study

For this study, we use the same set-up as in Chapter 4, that means we choose the following parameters: $\varphi = 0.55$, $\omega^2 = 0.6$, $\sigma^2 = 0.8$, and, again, e_t double-exponentially distributed and b_t normally distributed to simulate some realizations of the process $X_t = (\varphi + b_t)X_{t-1} + e_t$. As described before, the optimal bandwidths that minimize the *MSE* h and ε for the direct estimation of a density are of order $n^{-1/6}$. Since the e_t are double-exponentially distributed, we are in the smooth case. As we have seen, the optimal bandwidth k and δ for the deconvolution estimators are of order $\mathcal{O}\left(n^{-\frac{1}{2\beta+6}}\right)$. Fortunately, we can determine the constant β in advance because we have simulated data. We recall that we assumed the characteristic function to behave like $\frac{1}{\phi_e(u)} \leqslant c|u|^\beta \; \forall |u| \geqslant M$ with $c, \beta \geqslant 0$. For double-exponentially distributed random variables, the characteristic function is $\phi(t) = \frac{1}{1+\left(\frac{t}{a}\right)^2} \approx \frac{a^2}{t^2}$ for $t >> a$, thus $\beta = 2$ and the bandwidths should be of order $n^{-\frac{1}{10}}$. We will use these bandwidths but we will compare the estimators along their *MISE* since we want to have a global criterion in contrast to the local criterion *MSE* to judge if one whole estimated curve is better than another one.
Our approach is as follows: Given some simulated data, we want to estimate the density of the innovation noise on the interval $[-5, +5]$ with the estimator given in Definition 5.1.1 and the density of the disturbance noise with the estimator using the simple residuals as given in Definition 5.2.1 and with the the estimator using deconvolution techniques as given in Definition 5.3.3 as well as with the various variations of the estimator given in Definition 5.4.1 (multipoint estimator), Definition 5.4.4 (non-negative estimator), and Definition 5.4.6 (truncated estimator). All densities and characteristic functions that are continuous functions are discretized and considered on a very fine grid of step size $s = 0.02$.
First, we estimate the density of the innovation and the disturbance noise with the deconvolution technique with parameter a=1. For each simulated process we determine the estimated densities for various bandwidths and compare them by their *MISE*to decide which bandwidth is the optimal one for this process. This is repeated frequently for each value of n and for various values of n. We obtain $h = 4.7\; n^{-1/6}$, $\varepsilon = 4.6\; n^{-1/6}$, $k = 2.7\; n^{-1/10}$, $\delta = 4.7\; n^{-1/10}$ and refer to Figure 5.1 for further information.

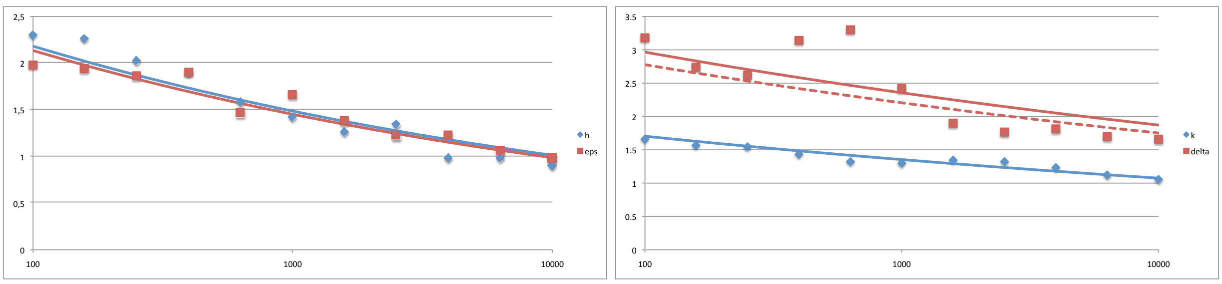

Figure 5.1: Optimal bandwidths for the estimator of the innovation noise (left) and the disturbance noise (deconvolution, right) and n varying between 100 and 10 000.

In a second step, we allow a to vary between 0.1 and 4.0 and proceed as before. For several values of a, we determine the optimal bandwidth depending on n and compute the regression parameter α_k, α_δ, for $k = \alpha_k n^{-1/10}$ and $\delta = \alpha_\delta n^{-1/10}$. The bandwidth k

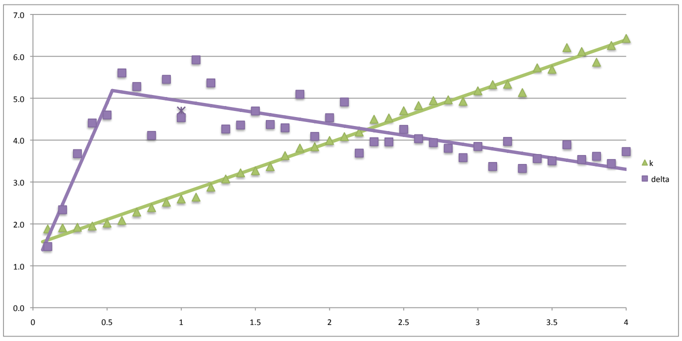

Figure 5.2: Regression parameters α_k and α_δ, for optimal bandwidths k and δ

should be increasing with a increasing while δ should be maximal for $a = 0.54$:

$$k = (1.49 + 1.23a)\, n^{-1/10}$$

$$\delta = \begin{cases} (0.86 + 8.07)\, n^{-1/10}, & a \leqslant 0.54 \\ (5.47 - 0.54a)\, n^{-1/10}, & a > 0.54 \end{cases}$$

We refer to Figure 5.2 for details. It shows the regression parameters α_k and α_δ, i.e. the displayed numbers have to be multiplied by $n^{-1/10}$ to obtain the actual bandwidth.
We used these bandwidths to perform the following simulations. All results cited are based on $N = 1\,000$ repetitions. Figure 5.3 shows the *MISE* of all introduced estimators on the basis of sample size $n = 100$ and $n = 1\,000$. It can be seen that the *MISE* decreases for increasing n. Using the simple residuals leads to a high variability in the estimators and tends to a larger *MISE* than the deconvolution methods. These can even reduce the error that is encountered by the estimation of the disturbance noise. This can be explained by the fact that the density of the innovation noise is tapered in zero what can hardly be rebuild by the estimator, see also Figure 5.5. However, the estimation leads to better results if the small residuals are used than if the large ones are used. The error of the estimator using deconvolution techniques is considerably smaller. Since for finite sample sizes the estimator will be negative as well, this error can be reduced further by the truncated and the non-negative estimator. However, this effect is small for small sample sizes and negliable for large sample sizes. For large sample sizes, the multi-point estimator has a considerable effect on the *MISE*. For small samples, the negative effect of estimating the density at points a that are very small or very large seems to dominate the positive effect of using more data points, but for larger sample sizes this is not the case anymore. As mentioned before, a good choice of a might be to choose it equal to one, and our simulations also indicated this, however, a slightly larger parameter $a \approx 1.4$ turns out to be even better in the simulations: The median, the upper and lower quartile as well as the upper and lower bound for outliers in the *MISE* between the estimators and the true density for various values of $a \in [-2.5, 2.5]$ are given in Figure 5.4 (right). The following graphs, Figure 5.5 through 5.11, show various types of estimators and estimators of various quality to obtain a sensation of the behavior of the estimators. For the estimation of the innovation noise and the deconvolution estimator, the concrete values of the *MISE* are displayed in Figure 5.4 (left), where it gets obvious that in few cases the error is small and that in most cases it is moderate but that there are also some cases where the error can be large and that there are some outliers as well.

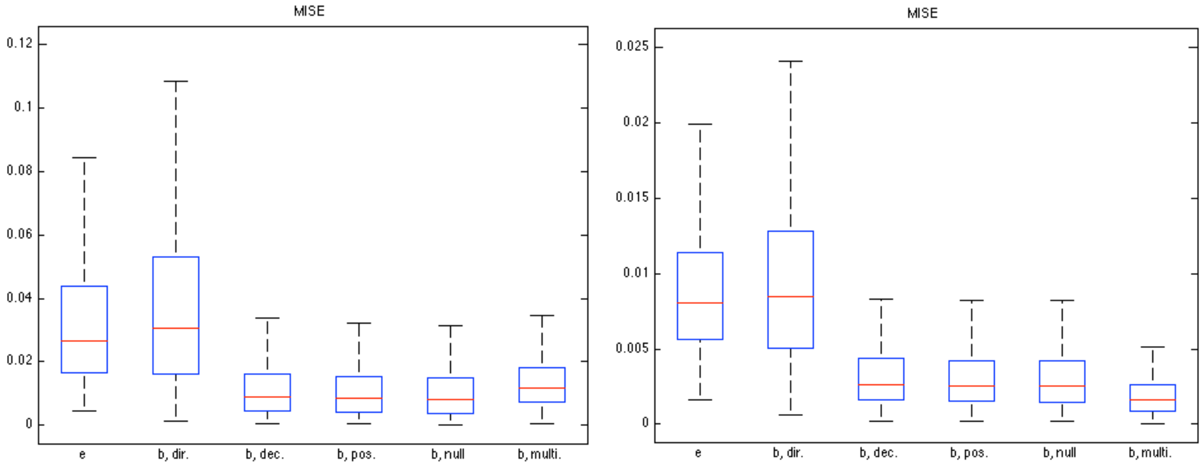

Figure 5.3: *MISE* for various estimation methods, $n = 100$ (left) and $n = 1\,000$ (right)

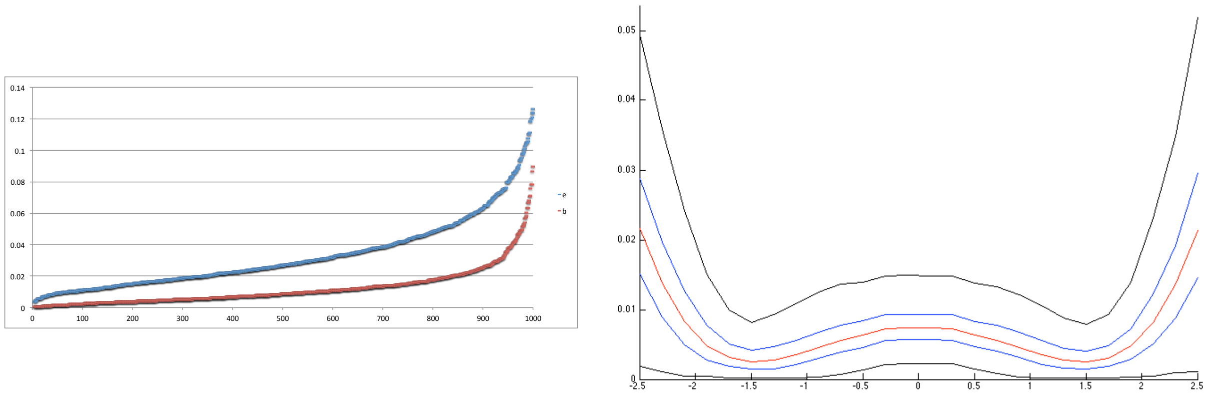

Figure 5.4: Estimated *MISE* of e_t and b_t (deconv.) with $n = 100$ (left) and *MISE* for various a with $n = 1000$ (right)

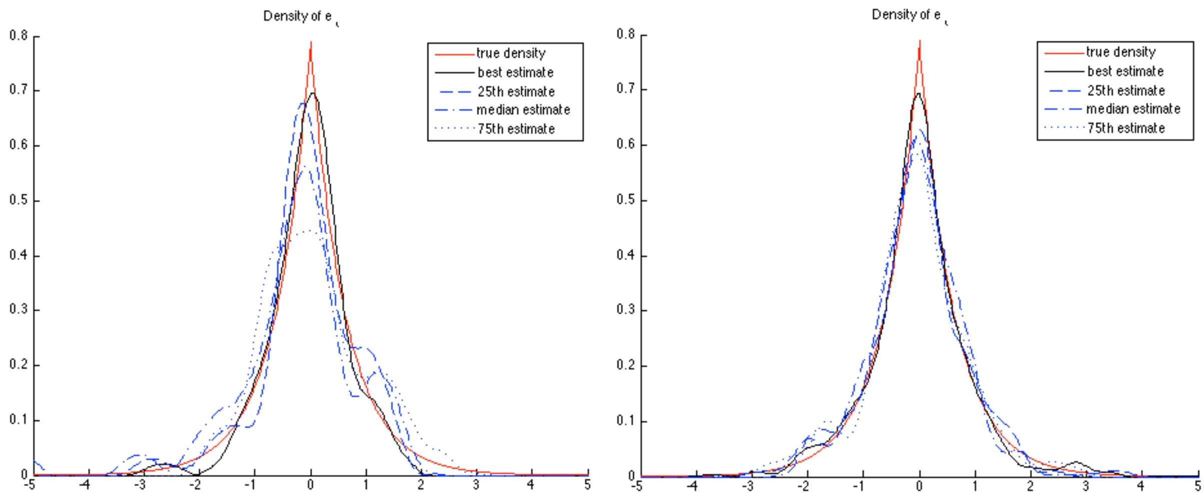

Figure 5.5: Estimated density of e_t with $n = 100$ (left) and $n = 1\,000$ (right)

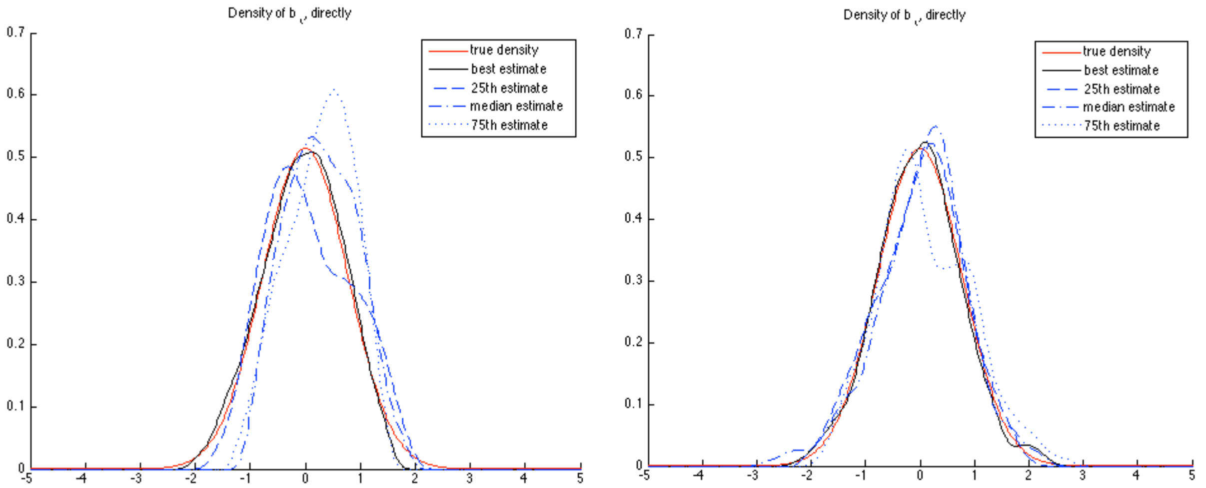

Figure 5.6: Estimated density of b_t (directly) with $n = 100$ (left) and $n = 1\,000$ (right)

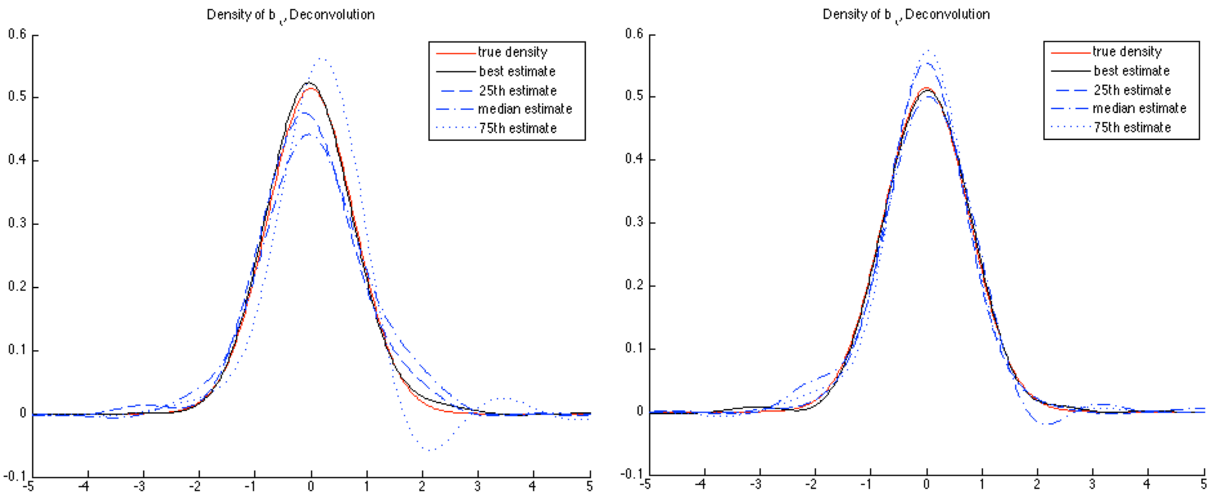

Figure 5.7: Est. density of b_t (deconvolution) with $n = 100$ (left) and $n = 1\,000$ (right)

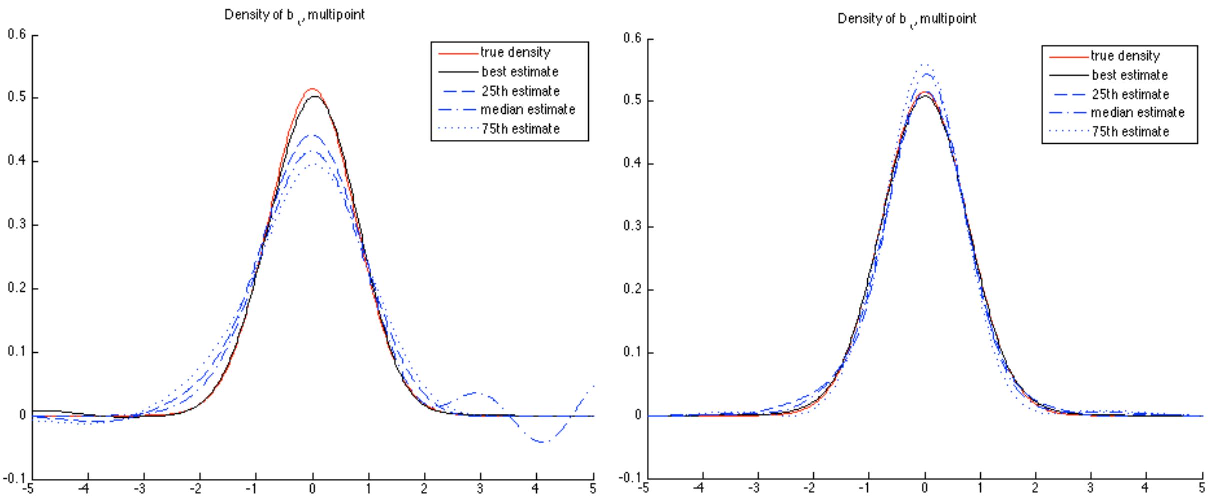

Figure 5.8: Estimated density of b_t (multipoint) with $n = 100$ (left) and $n = 1\,000$ (right)

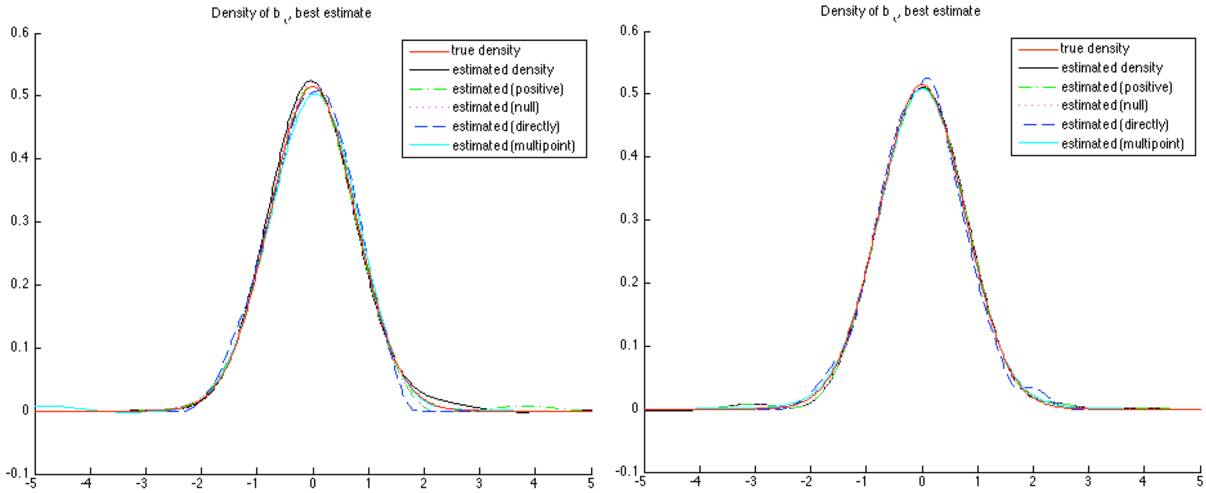

Figure 5.9: Est. dens. of b_t (best est., var. methods) with $n = 100$ (l) and $n = 1\,000$ (r)

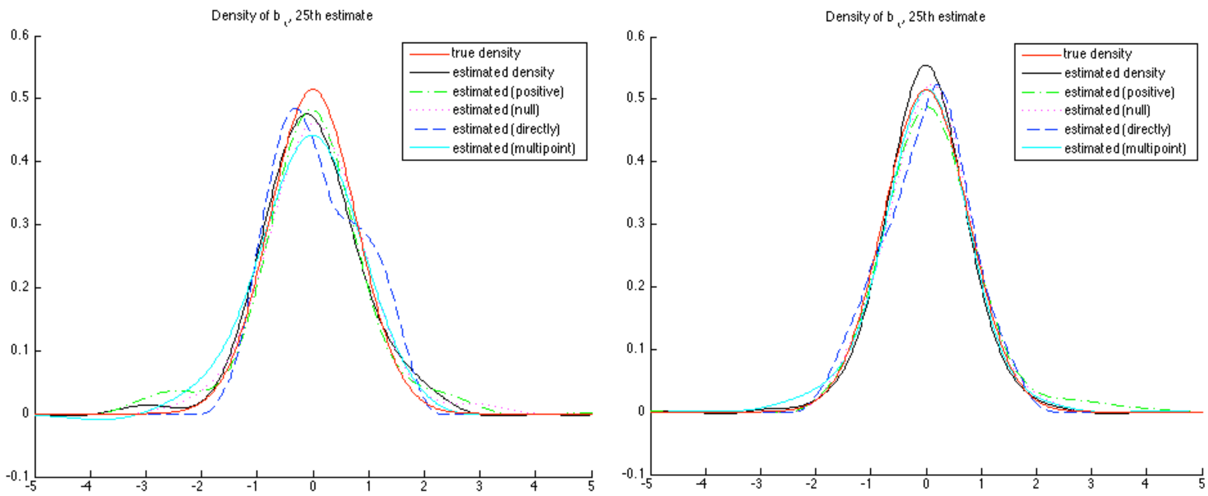

Figure 5.10: Est. dens. of b_t (25th est., var. methods) with $n = 100$ (l) and $n = 1\,000$ (r)

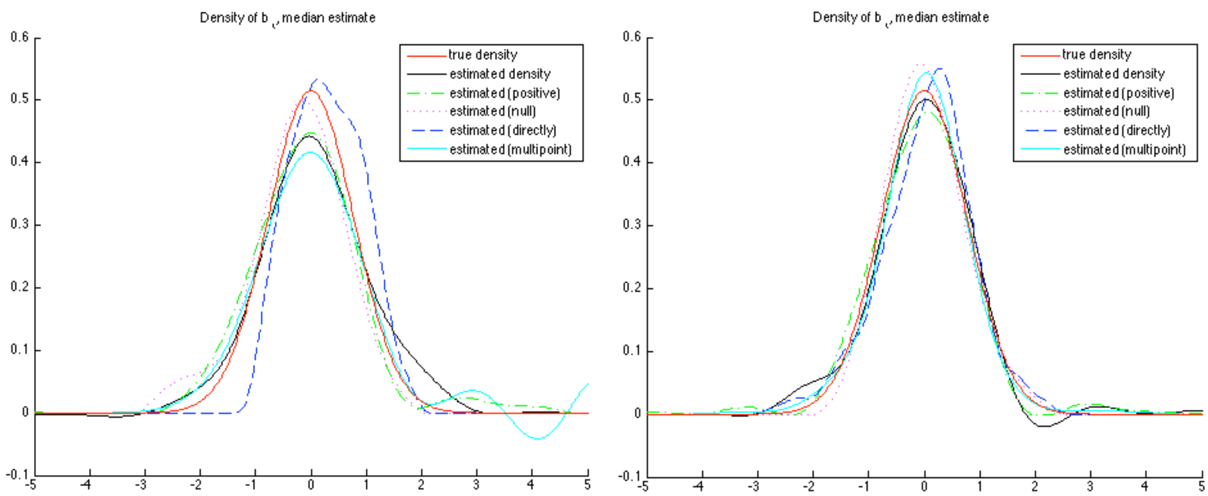

Figure 5.11: Est. dens. of b_t (med. est., var. methods) with $n = 100$ (l) and $n = 1\,000$ (r)

5.6 Proofs

In this section, we state the proofs that we have left out in the previous sections.

5.6.1 Innovation parameter

Consistency

We recall that in Definition 2.1.5 we defined the truncated RCA process to be

$$\tilde{X}_t^s = \sum_{i=0}^{s-1} \left(\prod_{j=0}^{i-1} (\varphi + b_{t-j}) \right) e_{t-i}. \tag{5.12}$$

Proof of Lemma 5.1.3. We note that we assume the second derivatives of f and k to be bounded. By taylor expansions for K and W first and for k and f thereafter, we obtain:

$$\begin{aligned} \mathbf{E}[A_1] &= \frac{1}{h\varepsilon} \int\int\int K\left(\frac{bx+e-y}{h}\right) W\left(\frac{x}{\varepsilon}\right) f(x)h(b)k(e)\mathrm{dxdbde} \\ &= \int\int\int K(u)W(v)f(0+v\varepsilon)h(b)k(hu+y-bv\varepsilon)\mathrm{dvdbdu} \\ &= k(y)f(0) + \mathcal{O}_P(h^2+\varepsilon^2) \end{aligned} \tag{5.13}$$

While the expectation of the estimator could be determined easily, we have to state some considerations first to determine its variance. To bound the sum of autocovariances

$$\sum_{t=2}^{\infty} \mathbf{Cov}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right) W\left(\frac{X_0}{\varepsilon}\right), K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]$$

we note that we can subtract the truncated process $\tilde{X}_{t-1}^{t-1}$ from X_{t-1} because it is independent of X_0 so that the term is equal to

$$\begin{aligned} &\sum_{t=2}^{\infty} \mathbf{Cov}\Bigg[K\left(\frac{b_1X_0+e_1-y}{h}\right) W\left(\frac{X_0}{\varepsilon}\right), \\ &\qquad \left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) - K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right) W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)\Bigg]. \end{aligned}$$

For the first summand we have

$$\begin{aligned} &\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right) W\left(\frac{X_0}{\varepsilon}\right) K\left(\frac{b_2X_1+e_2-y}{h}\right) W\left(\frac{X_1}{\varepsilon}\right)\right] \\ &\quad \leqslant \mathbf{E}\left[hK\left(\frac{b_1X_0+e_1-y}{h}\right) W\left(\frac{X_0}{\varepsilon}\right) \int K(v)\,\mathrm{dv} W\left(\frac{(\varphi+b_1)X_0+e_1}{\varepsilon}\right) k(\Theta)\mathrm{du}\right] \\ &\quad = \mathbf{E}\left[h\varepsilon \int K\left(\frac{b_1X_0+\varepsilon v+(\varphi+b_1)X_0-y}{h}\right) W\left(\frac{X_0}{\varepsilon}\right) W(v)\, k(\Theta)\mathrm{dv}\right] \\ &\quad \leqslant \mathbf{E}\left[h\varepsilon \int K(\Xi)\, W\left(\frac{X_0}{\varepsilon}\right) W(v)\, k(\Theta)\mathrm{dv}\right] \\ &\quad \leqslant h\varepsilon^2 C \end{aligned}$$

To evaluate the other summands we introduce a suitable $M = M(n) \overset{n\to\infty}{\longrightarrow} \infty$ such that $Mh^2\varepsilon^2 \overset{n\to\infty}{\longrightarrow} 0$ and $\alpha^M \overset{n\to\infty}{\longrightarrow} 0$ for $|\alpha| < 1$. Then we can split up the sum into the first M-2 summands and the residual sum. For the first summands we obtain:

$$
\begin{aligned}
&\sum_{t=3}^{M} \mathbf{Cov}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right),\right.\\
&\qquad\left.\left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)-K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)\right]\\
&\quad=\sum_{t=3}^{M}\left\{\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right.\right.\\
&\qquad\left.\left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)-K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)\right]\\
&\qquad-\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]\\
&\qquad\left.\cdot\,\mathbf{E}\left[K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)-K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right]\right\}\\
&\quad=\mathcal{O}\left(Mh^2\varepsilon^2\right)
\end{aligned}
$$

by using the following calculation for the first part and a taylor expansion and substitution for the second part.

$$
\begin{aligned}
&\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\\
&\quad=h\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\int K(v)k(hv+y-b_tX_{t-1})\mathrm{dv}\right]\\
&\quad\leqslant\mathbf{E}\left[hK\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\int K(v)k(\Theta)\mathrm{dv}\right]\\
&\quad=\mathbf{E}\left[hK\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\int W\left(\frac{(\varphi+b_{t-1})X_{t-2}+u}{\varepsilon}\right)k(u)\mathrm{du}k(\Theta)\right]\\
&\quad=h\varepsilon\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\int W(v)k(\varepsilon v-(\varphi+b_{t-1})X_{t-2})\mathrm{dv}k(\Theta)\right]\\
&\quad\leqslant h\varepsilon\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\int W(v)k(\Theta)\mathrm{dv}k(\Theta)\right]\\
&\quad=h^2\varepsilon^2C
\end{aligned}
$$

Before we can consider the residual sum, we have to state some additional thoughts. By direct computations and taylor expansions as in Equation (5.13) and substitutions we obtain for $c \geqslant 0$

$$
\mathbf{E}\left[\frac{1}{h\varepsilon}K^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right]=k(y)f(0)c+\mathcal{O}(h+\varepsilon).
$$

Further, if $H(a,b,c) = K\left(a - \varphi b - \frac{y}{h}\right)W(c)$ is Lipschitz continuous with constant L_H, a geometrically decaying bound for a term of a squared expectation is given by:

$$\begin{aligned}
&\mathbf{E}\left[\left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) - K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&\quad= \mathbf{E}\left[\left(K\left(\frac{X_t-\varphi X_{t-1}-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) - K\left(\frac{\tilde{X}_t^{t-1}-\varphi\tilde{X}_{t-1}^{t-1}-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&\quad= \mathbf{E}\left[\left(H\left(\frac{X_t}{h},\frac{X_{t-1}}{h},\frac{X_{t-1}}{\varepsilon}\right) - H\left(\frac{\tilde{X}_t^{t-1}}{h},\frac{\tilde{X}_{t-1}^{t-1}}{h},\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&\quad\leqslant L_H\mathbf{E}\left[\left(\frac{X_t-\tilde{X}_t^{t-1}}{h}\right)^2 + \left(\frac{X_{t-1}-\tilde{X}_{t-1}^{t-1}}{h}\right)^2 + \left(\frac{X_{t-1}-\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)^2\right]^{\frac{1}{2}} \\
&\quad= L_H\sqrt{\left(\frac{2}{h^2}+\frac{1}{\varepsilon^2}\right)\frac{\sigma^2}{1-\vartheta}\vartheta^{t-1}} \\
&\quad\leqslant \sqrt{\vartheta^{t-1}}L_H\frac{h+\varepsilon}{h\varepsilon}\sqrt{\frac{2\sigma^2}{1-\vartheta}}
\end{aligned}$$

If $H^*(a,b,c) = K'\left(a - \varphi b - \frac{y}{h}\right)W(c)$ is Lipschitz continuous and bounded with Lipschitz constant $\tilde{L}_H$, we note that the function $X_tH^*(\frac{X_{t+1}}{h},\frac{X_t}{h},\frac{X_t}{\varepsilon})$ is Lipschitz continuous as well (write $\hat{X}_t = \left(\frac{X_{t+1}}{h},\frac{X_t}{h},\frac{X_t}{\varepsilon}\right)$):

$$\begin{aligned}
||X_tH^*(\hat{X}_t) - Y_tH^*(\hat{Y}_t)|| &\leqslant ||X_tH^*(\hat{X}_t) - X_tH^*(\hat{Y}_t)|| + ||X_tH^*(\hat{Y}_t) - Y_t|H^*(\hat{Y}_t)|| \\
&= ||X_t||\,||H^*(\hat{X}_t) - H^*(\hat{Y}_t)|| + ||X_t - Y_t||\,||H^*(\hat{Y}_t)|| \\
&\leqslant ||X_t||L_{H*}||X_t - Y_t|| + \kappa||X_t - Y_t|| \\
&= \left(\sqrt{\mathbf{E}\left[X_t^2\right]}L_{H*} + \kappa\right)||X_t - Y_t||
\end{aligned}$$

so that a similar computation as before yields that

$$\mathbf{E}\left[\left(X_{t-1}K'\left(\Xi'\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) - \tilde{X}_{t-1}^{t-1}K'\left(\Xi'\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \leqslant \sqrt{\vartheta^{t-1}}\tilde{L}_H\frac{1}{\varepsilon}\sqrt{\frac{2\sigma^2}{1-\vartheta}}.$$

Now, we can consider the residual sum of the autocovariances and obtain by first applying the Cauchy-Schwarz-Inequality and the computations above thereafter:

$$\begin{aligned}
&\sum_{t=M+1}^{\infty}\mathbf{Cov}\Bigg[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right), \\
&\qquad\left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) - K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sum_{t=M+1}^{\infty} \mathbf{E}\left[\left(K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&\quad \mathbf{E}\left[\left(K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)-K\left(\frac{b_t\tilde{X}_{t-1}^{t-1}+e_t-y}{h}\right)W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&\leqslant \sum_{t=M+1}^{\infty} C_1\sqrt{h\varepsilon}\sqrt{\vartheta^{t-1}}L_H\frac{h+\varepsilon}{h\varepsilon}\sqrt{\frac{2\sigma^2}{1-\vartheta}} \\
&= C_1L_H\frac{h+\varepsilon}{\sqrt{h\varepsilon}}\sqrt{\frac{2\sigma^2}{1-\vartheta}}\sqrt{\vartheta^M}\sum_{t=0}^{\infty}\sqrt{\vartheta^t} \\
&= \sqrt{\vartheta^M}\frac{h+\varepsilon}{\sqrt{h\varepsilon}}C_1L_H\sqrt{\frac{2\sigma^2}{1-\vartheta}}\frac{1}{1-\sqrt{\vartheta}}
\end{aligned}
$$

With these preliminaries we can directly determine the variance of the estimator:

$$
\begin{aligned}
\mathbf{Var}[A_1] &= \frac{1}{n^2h^2\varepsilon^2}\mathbf{Var}\left[\sum_{t=1}^{n}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] \\
&\leqslant \left|\frac{1}{nh\varepsilon}\mathbf{E}\left[\frac{1}{h\varepsilon}K^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right]\right| \\
&\qquad + \frac{2}{n^2h^2\varepsilon^2}\sum_{i=1}^{n}\underbrace{(n-i+1)}_{\leqslant n}\left|\mathbf{Cov}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right),\right.\right. \\
&\qquad\qquad \left.\left. K\left(\frac{b_iX_{i-1}+e_i-y}{h}\right)W\left(\frac{X_{i-1}}{\varepsilon}\right)\right]\right| \\
&\leqslant \frac{1}{nh\varepsilon}\left(k(y)f(0)c+\mathcal{O}(h)+\mathcal{O}(\varepsilon)\right) \\
&\qquad + \frac{1}{nh^2\varepsilon^2}\left(Mh^2\varepsilon^2C_1+\alpha^M(1-\alpha^{n-M})\frac{\varepsilon+h}{\sqrt{h\varepsilon}}C_2^*+h\varepsilon^2C_3\right) \\
&\leqslant \frac{1}{nh\varepsilon}\left(k(y)f(0)c+\mathcal{O}(h)+\mathcal{O}(\varepsilon)\right)+\frac{M}{n}C_1+\frac{\alpha^M(h+\varepsilon)}{nh^{2.5}\varepsilon^{2.5}}C_2^*+\frac{1}{nh}C_3 \\
&= \mathcal{O}\left(\frac{1}{nh\varepsilon}\right)+o\left(\frac{1}{nh\varepsilon}\right)
\end{aligned}
$$

This completes the proof of the Lemma. □

Proof of Lemma 5.1.4. By taylor expansion and substitution we obtain

$$
\mathbf{E}[A_2] = \int W(u)f(0)\mathrm{d}u + \int W(u)f'(0)u\varepsilon\mathrm{d}u + \int \varepsilon^2u^2W(u)f''(\zeta u\varepsilon)\mathrm{d}u = f(0)+\mathcal{O}\left(\varepsilon^2\right)
$$

$$
\begin{aligned}
\mathbf{Var}[A_2] &\leqslant \frac{1}{n\varepsilon^2}\mathbf{E}\left[W^2\left(\frac{X_0}{\varepsilon}\right)\right]+\frac{2}{n\varepsilon^2}\sum_{t=2}^{n}\left|\mathbf{Cov}\left[W\left(\frac{X_0}{\varepsilon}\right),W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\right| \\
&= \frac{1}{n\varepsilon^2}\int W^2\left(\frac{x}{\varepsilon}\right)f(x)\mathrm{d}x+\frac{2}{n\varepsilon^2}\left(M\varepsilon C_1+\alpha^M(1-\alpha^{n-M})\sqrt{\varepsilon}C_2+\varepsilon C_3\right)=\mathcal{O}\left(\frac{1}{n\varepsilon}\right).
\end{aligned}
$$

□

Proof of Lemma 5.1.5. As we will see soon, we have that

$$\frac{1}{nh\varepsilon}\sum_{t=1}^{n} X_{t-1}K'\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) = \mathcal{O}_P\left(h\varepsilon^2+\varepsilon^3+\frac{\sqrt{\varepsilon}}{\sqrt{nh}}\right).$$

Additionally, we assumed that a $\sqrt{n}$-consistent estimator $\hat{\varphi}$ for φ is given, so that

$$B_1 = \frac{\varphi-\hat{\varphi}}{h}\frac{1}{nh\varepsilon}\sum_{t=1}^{n} X_{t-1}K'\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) = \mathcal{O}_P\left(\frac{\varepsilon^2}{\sqrt{n}}+\frac{\varepsilon^3}{h\sqrt{n}}+\frac{\sqrt{\varepsilon}}{nh^{1.5}}\right).$$

The first assertion can be seen as follows:

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{h\varepsilon}X_{t-1}K'\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] \\
&\qquad = \int\!\!\int\!\!\int \frac{1}{h\varepsilon}xK'\left(\frac{bx+e-y}{h}\right)W\left(\frac{x}{\varepsilon}\right)f(x)k(e)h(b)\mathrm{dxdbde} = \mathcal{O}\left(h\varepsilon^2+\varepsilon^3\right)
\end{aligned}$$

In addition,

$$\begin{aligned}
&\frac{1}{n^2h^2\varepsilon^2}\mathbf{Var}\left[\sum_{t=1}^{n} X_{t-1}K'\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] \\
&\qquad \leqslant \left|\frac{1}{nh\varepsilon}\mathbf{E}\left[\frac{1}{h\varepsilon}X_0^2K'^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right]\right| \\
&\qquad\quad + \frac{2}{nh^2\varepsilon^2}\left(\sum_{i=3}^{n}\left|\mathbf{Cov}\left[X_0K'\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right),\right.\right.\right. \\
&\qquad\qquad\qquad \left.\left.\left. X_{i-1}K'\left(\frac{b_iX_{i-1}+e_i-y}{h}\right)W\left(\frac{X_{i-1}}{\varepsilon}\right)\right]\right| + h\varepsilon^2C^*\right) \\
&\qquad = \mathcal{O}\left(\frac{\varepsilon}{nh}+\frac{M}{n}+\alpha^M\left(\frac{1}{nh^2}+\frac{1}{nh\varepsilon}\right)+\frac{\varepsilon}{nh}\right)
\end{aligned}$$

since

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{h\varepsilon}X_0^2K'^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right] \\
&\qquad = \int\!\!\int\!\!\int K^2(u)W^2(v)v^2f(0+v\varepsilon)h(b)k(hu+y-bv\varepsilon)\mathrm{dbdvdu} \\
&\qquad = \varepsilon^2\left(k(y)f(0)+\mathcal{O}(h)+\mathcal{O}(\varepsilon)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=3}^{n}\left|\mathbf{Cov}\left[X_0K'\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right),\right.\right. \\
&\qquad\qquad\qquad \left.\left. X_{t-1}K'\left(\frac{b_iX_{i-1}+e_i-y}{h}\right)W\left(\frac{X_{i-1}}{\varepsilon}\right)\right]\right| \\
&= \mathcal{O}\left(Mh^2\varepsilon^2+\varepsilon\alpha^M(\varepsilon+h)\right)
\end{aligned}$$

by the argumentation and computations of the proof of Lemma 5.1.3 and

$$\mathbf{E}\left[X_0 W\left(\frac{X_0}{\varepsilon}\right)\right] = \int xW\left(\frac{x}{\varepsilon}\right) f(x)\mathrm{dx} = \varepsilon \int u\varepsilon W(u) f(u\varepsilon)\mathrm{du} = \mathcal{O}(\varepsilon^2)$$

□

We conclude this part with the

Proof of Lemma 5.1.6. We first note that

$$\begin{aligned} 0 &\leqslant \frac{1}{h\varepsilon} X_{t-1}^2 K''\left(\frac{b_t X_{t-1} + e_t - y}{h} + \vartheta \frac{\varphi - \hat{\varphi}}{h} X_{t-1}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \\ &\leqslant \frac{1}{h\varepsilon} X_{t-1}^2 K''\left(\Xi''\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) \end{aligned}$$

and will see soon that

$$\frac{1}{nh\varepsilon}\sum_{t=1}^{n} X_{t-1}^2 K''\left(\frac{b_t X_{t-1} + e_t - y}{h} + \vartheta \frac{\varphi - \hat{\varphi}}{h} X_{t-1}\right) = \mathcal{O}_P\left(\frac{\varepsilon^2}{h} + \frac{\varepsilon^{1.5}}{h\sqrt{n}}\right)$$

so that with the assumption that a $\sqrt{n}$-consistent estimator for φ is given we have

$$B_2 = \mathcal{O}_P\left(\frac{1}{h^2 n}\right) \cdot \mathcal{O}_P\left(\frac{\varepsilon^2}{h} + \frac{\varepsilon^{1.5}}{h\sqrt{n}}\right) = \mathcal{O}_P\left(\frac{\varepsilon^{1.5}}{n^{1.5}h^3}\right).$$

The first assertion can be seen as follows:

$$\mathbf{E}\left[\frac{1}{h\varepsilon} X_{t-1}^2 K''\left(\Xi''\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] = \frac{1}{h} K''(\Xi'') \int v^2\varepsilon^2 W(v) f(v\varepsilon)\mathrm{dv} = \mathcal{O}\left(\frac{\varepsilon^2}{h}\right)$$

and

$$\begin{aligned} &\frac{1}{n^2h^2\varepsilon^2}\mathbf{Var}\left[\sum_{t=1}^{n} X_{t-1}^2 W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] \\ &\leqslant \left|\frac{1}{nh\varepsilon}\mathbf{E}\left[\frac{1}{h\varepsilon} X_0^4 W^2\left(\frac{X_0}{\varepsilon}\right)\right]\right| + \frac{2}{nh^2\varepsilon^2}\left(\sum_{i=3}^{n}\left|\mathbf{Cov}\left[X_0^2 W\left(\frac{X_0}{\varepsilon}\right), X_{i-1}^2 W\left(\frac{X_{i-1}}{\varepsilon}\right)\right]\right| + \varepsilon^5 C^*\right) \\ &= \frac{1}{nh\varepsilon}\mathcal{O}\left(\frac{\varepsilon^4}{h}\right) + \frac{2}{nh^2\varepsilon^2}\mathcal{O}\left(M\varepsilon^6 + \varepsilon^2\alpha^M + \varepsilon^5\right) + \frac{2C^*\varepsilon^3}{nh^2} \\ &= \mathcal{O}\left(\frac{\varepsilon^3}{nh^2} + \frac{M\varepsilon^4}{nh^2} + \alpha^M\frac{1}{nh^2}\right) \end{aligned}$$

since

$$\mathbf{E}\left[\frac{1}{h\varepsilon} X_0^4 W^2\left(\frac{X_0}{\varepsilon}\right)\right] = \mathcal{O}\left(\frac{\varepsilon^4}{h}\right)$$

and

$$
\begin{aligned}
&\sum_{i=3}^{n}\left|\mathbf{Cov}\left[X_0^2 W\left(\frac{X_0}{\varepsilon}\right), X_{t-1}^2 W\left(\frac{X_{i-1}}{\varepsilon}\right)\right]\right| \\
&\leqslant \sum_{t=3}^{M}\mathbf{E}\left[X_0^2 W\left(\frac{X_0}{\varepsilon}\right) X_{t-1}^2 W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - \sum_{t=3}^{M}\mathbf{E}\left[X_0^2 W\left(\frac{X_0}{\varepsilon}\right)\left(\tilde{X}_{t-1}^{t-1}\right)^2 W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right] \\
&\quad + \sum_{t=M+1}^{\infty}\mathbf{E}\left[X_0^4 W\left(\frac{X_0}{\varepsilon}\right)\right]^{\frac{1}{2}} \cdot \mathbf{E}\left[\left(X_{t-1}^2 W\left(\frac{X_{t-1}}{\varepsilon}\right) - \left(\tilde{X}_{t-1}^{t-1}\right)^2 W\left(\frac{\tilde{X}_{t-1}^{t-1}}{\varepsilon}\right)\right)^2\right]^{\frac{1}{2}} \\
&= \mathcal{O}\left(M\varepsilon^6 + \alpha^M\varepsilon^2\right)
\end{aligned}
$$

by the argumentation of the proof of Lemmas 5.1.3 and 5.1.5. □

Thus, we have shown the asymptotic consistency of the estimator.

Asymptotic distribution

Proof of Theorem 5.1.8. We recall that by taylor expansion the estimator $\hat{k}(y)$ splits up into three parts (c.f. Equation (5.1)) and define the random variables $Z_{n,t}(y)$, $t = 1, \dots, n$, $n = 1, 2, \dots$ with $\mathbf{E}[Z_{n,t}(y)] = 0$ by

$$
\begin{aligned}
&Z_{n,t}(y) \\
&= \frac{1}{\sqrt{h\varepsilon}}\left\{K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right) - \mathbf{E}\left[K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\right\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\sqrt{nh\varepsilon}\left(\frac{A_1}{A_2} - k(y)\right) \\
&\quad = \sqrt{nh\varepsilon}\left(\frac{\frac{1}{nh\varepsilon}\sum_{t=1}^{n} K\left(\frac{b_t X_{t-1}+e_t-y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)}{\frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)} - k(y)\right) \\
&\quad = \frac{1}{\frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sqrt{h\varepsilon}}\left\{K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right.\right. \\
&\qquad\qquad \left. - \mathbf{E}\left[K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\right\} \\
&\qquad \left. + \sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\}\right) \\
&\quad = \frac{1}{\frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{n,t}(y)\right. \\
&\qquad \left. + \sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_t X_{t-1} + e_t - y}{h}\right) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y) W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\}\right)
\end{aligned}
$$

Now, we can consider each term separately.

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{n,t}(y) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma^2) \quad i.D., \quad \text{see Theorem 5.6.2,}\\
&\frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right) \overset{n\to\infty}{\longrightarrow} f(0) > 0 \quad i.P., \quad \text{see Lemma 5.1.4,}\\
&\sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\}\\
&\qquad\qquad \overset{n\to\infty}{\longrightarrow} B \quad i.P., \quad \text{see Lemma 5.6.1}
\end{aligned}
$$

This yields the proposition for the main part by applying the Theorem of Slutsky. Furthermore, it directly follows from Lemmas 5.1.5 and 5.1.6 that the other two terms resulting out of the kernel's taylor expansion (Equation (5.1)) vanish asymptotically:

$$
\frac{\varphi-\hat{\varphi}}{h}\frac{1}{\sqrt{nh\varepsilon}}\sum_{t=1}^{n} X_{t-1}K'\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right) = \mathcal{O}_P\left(\frac{\varepsilon^{2.5}h^{0.5}}{n^{0.5}}+\frac{\varepsilon^{3.5}}{h^{0.5}}+\frac{\varepsilon}{hn^{0.5}}\right)
$$

$$
\begin{aligned}
\left(\frac{\varphi-\hat{\varphi}}{h}\right)^2 \frac{1}{\sqrt{nh\varepsilon}}\sum_{t=1}^{n} X_{t-1}^2 K''\left(\frac{b_tX_{t-1}+e_t-y}{h}+\vartheta\frac{\varphi-\hat{\varphi}}{h}\right. &\left.X_{t-1}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\\
&= \mathcal{O}_P\left(\varepsilon^{2.5}\sqrt{h}+\frac{\varepsilon^{3.5}}{\sqrt{h}}+\frac{\varepsilon}{h\sqrt{n}}\right)
\end{aligned}
$$

□

We now show the results we just cited.

Lemma 5.6.1 (Determination of the Bias)**.** *For the bias B holds true:*

$$
\sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\boldsymbol{E}\left[\frac{1}{h}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\} \overset{n\to\infty}{\longrightarrow} B \quad i.P.
$$

with B given in Theorem 5.1.8.

Proof. By Lemmas 5.1.3 and 5.1.4

$$
\begin{aligned}
&\mathbf{E}\left[\sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\}\right]\\
&= \sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - \mathbf{E}\left[k(y)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\right\}\\
&= \sqrt{nh\varepsilon}\left(K_1h^2+K_2\varepsilon^2+K_3h\varepsilon+\mathcal{O}\left(h^3+\varepsilon^3\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbf{Var}\left[\sqrt{nh\varepsilon}\frac{1}{n\varepsilon}\sum_{t=1}^{n}\left\{\mathbf{E}\left[\frac{1}{h}K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right] - k(y)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right\}\right]\\
&= k^2(y)nh\varepsilon\mathbf{Var}\left[\frac{1}{n\varepsilon}\sum_{t=1}^{n} W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= k^2(y)nh\varepsilon\,\mathcal{O}\left(\frac{1}{n\varepsilon}\right)\\
&= \mathcal{O}\left(h\right)
\end{aligned}
$$

□

Theorem 5.6.2 (CLT for the $Z_{n,t}(y)$)**.** *With σ^2 as given in Theorem 5.1.8 it holds that*

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{n,t}(y) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma^2) \quad i.D.
$$

Proof. The central limit theorem for weak dependent random variables (Neumann & Paparoditis (2008), Th. 6.1) yields the desired convergence if we can show that the pre-requisites are met.

Variance: Using Lemma 5.1.3 we obtain

$$
\begin{aligned}
&\mathbf{E}[Z_{n,1}^2]\\
&= \frac{1}{h\varepsilon}\mathbf{E}\left[\left(K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)-\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]\right)^2\right]\\
&= \frac{1}{h\varepsilon}\left(\mathbf{E}\left[K^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right]-\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]^2\right)\\
&= k(y)f(0)\int K^2(u)\mathrm{du}\int W^2(v)\mathrm{dv}+\mathcal{O}\left(h^2+\varepsilon^2\right)-\frac{\left(h\varepsilon\left(k(y)f(0)+\mathcal{O}\left(h^2+\varepsilon^2\right)\right)\right)^2}{h\varepsilon}\\
&= k(y)f(0)\int K^2(u)\mathrm{du}\int W^2(v)\mathrm{dv}+\mathcal{O}\left(h^2+\varepsilon^2\right)
\end{aligned}
$$

Autocovariances:

$$
\begin{aligned}
&\sum_{t=2}^{\infty}\mathbf{E}[Z_{n,1}(y)Z_{n,t}(y)]\\
&\qquad= \frac{1}{h\varepsilon}\sum_{t=2}^{\infty}\mathbf{Cov}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right),K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\\
&\qquad\leqslant Mh\varepsilon C_1+\alpha^M\frac{\varepsilon+h}{(h\varepsilon)^{\frac{3}{2}}}C_2+\varepsilon C_3
\end{aligned}
$$

by the argumentation of the proof of Lemma 5.1.3 so that

$$
\sigma_n^2=\sum_{t\in\mathbb{Z}}\mathbf{E}\left[Z_{n,1}(y)Z_{n,t}(y)\right]=k(y)f(0)\int K^2(u)\mathrm{du}\int W^2(v)\mathrm{dv}+\mathcal{O}\left(h^2+\varepsilon\right).
$$

Lindeberg Condition: Let $\xi > 0$ and note that we assume the kernel functions to have finite fourth moments. We then obtain by applying the Cauchy Schwary Inequality

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\mathbf{E}\left[Z_{n,t}(y)^2\mathbb{1}_{\{|Z_{n,t}(y)|\geqslant\xi\sqrt{n}\}}\right]\\
&\quad=\frac{1}{n}\sum_{t=1}^{n}\mathbf{E}\Bigg[\Bigg(\frac{1}{\sqrt{h\varepsilon}}\Bigg\{K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\\
&\qquad\qquad-\mathbf{E}\left[K\left(\frac{b_tX_{t-1}+e_t-y}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\right]\Bigg\}\Bigg)^2\mathbb{1}_{\{|Z_{n,t}(y)|\geqslant\xi\sqrt{n}\}}\Bigg]\\
&\quad\leqslant\mathbf{E}\Bigg[\frac{1}{h^2\varepsilon^2}\Bigg\{K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\\
&\qquad\qquad-\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]\Bigg\}^4\Bigg]^{\frac{1}{2}}\mathbf{E}\left[\mathbb{1}_{\{|Z_{n,1}(y)|\geqslant\xi\sqrt{n}\}}\right]^{\frac{1}{2}}\\
&\quad=\frac{1}{h^2\varepsilon^2}\Bigg(\mathbf{E}\left[K^4\left(\frac{b_1X_0+e_1-y}{h}\right)W^4\left(\frac{X_0}{\varepsilon}\right)\right]\\
&\qquad-3\mathbf{E}\left[K^3\left(\frac{b_1X_0+e_1-y}{h}\right)W^3\left(\frac{X_0}{\varepsilon}\right)\right]\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]\\
&\qquad+6\mathbf{E}\left[K^2\left(\frac{b_1X_0+e_1-y}{h}\right)W^2\left(\frac{X_0}{\varepsilon}\right)\right]\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]^2\\
&\qquad-3\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]^3\\
&\qquad+\mathbf{E}\left[K\left(\frac{b_1X_0+e_1-y}{h}\right)W\left(\frac{X_0}{\varepsilon}\right)\right]^4\Bigg)^{\frac{1}{2}}P\left\{|Z_{n,1}(y)|\geqslant\xi\sqrt{n}\right\}^{\frac{1}{2}}\\
&\quad\leqslant\mathcal{O}\left(\left(\frac{1}{h^2\varepsilon^2}\left(\varepsilon h+\varepsilon^2h^2+\varepsilon^3h^3+\varepsilon^4h^4\right)\right)^{\frac{1}{2}}\right)\frac{1}{\xi\sqrt{n}}\mathbf{E}\left[Z_{n,1}^2(y)\right]^{\frac{1}{2}}\\
&\quad=\frac{1}{\sqrt{nh\varepsilon}}\frac{1}{\xi}\left(k(y)f(0)\int K^2(u)\mathrm{d}u\int W^2(v)\mathrm{d}v+\mathcal{O}\left(h^2+\varepsilon^2\right)\right)\quad\text{(see above)}
\end{aligned}
$$

Weak-dependance condition I: Let f be measurable and squere-integrable and $s_1 < \cdots < s_u < s_u + r = t_1 \leqslant t_2$. Note that we can write $Z_{n,t}$ as a function g of $X_{n,t}$ and $X_{n,t-1}$. By assumption and the argumentation about V in the proof of Lemma 5.1.3 this function g is Lipschitz continuous since it is a function of the kernels K and W. Further, we define for an arbitrary function f the function h by $h = f \circ g$ and recall that in Definition 2.1.5 for the process X_t the truncated version $\tilde{X}_t^s$ was defined to obtain

$$
\begin{aligned}
&\left|\mathbf{Cov}\left[f\left(Z_{n,s_1},\ldots,Z_{n,s_n}\right),Z_{n,t_1}\right]\right|\\
&\quad=\left|\mathbf{Cov}\left[f\left(g(X_{n,s_1},X_{n,s_1-1}),\ldots,g(X_{n,s_n},X_{n,s_n-1})\right),g(X_{n,t_1},X_{n,t_1-1})\right]\right|\\
&\quad=\left|\mathbf{Cov}\left[h\left(X_{n,s_1-1},X_{n,s_1},\ldots,X_{n,s_u-1},X_{n,s_u}\right),g(X_{n,t_1},X_{n,t_1-1})\right]\right|,\quad\text{for suitable } h
\end{aligned}
$$

renumber: $s_1 - 1 = s'_1 < s'_2 < \dots < s'_v = s_u \leqslant s'_v + r' = t'_1 < t'_2 = t_1$
with $r' = r - 1$ to obtain for $r \geqslant 2$ (note that then $s'_v < t'_1$) :

$$
\begin{aligned}
&= \left|\mathbf{Cov}\left[h\left(X_{n,s'_1}, X_{n,s'_2}, \dots, X_{n,s'_v}\right), g(X_{n,t'_1}, X_{n,t'_2})\right]\right| \\
&\leqslant \left(\mathbf{E}\left[h^2(X_{n,s'_1}, \dots, X_{n,s'_v})\right]\right)^{\frac{1}{2}} \cdot L \cdot \mathbf{E}\left[\left(X_{t_1} - \tilde{X}_{t_1}\right)^2 + \left(X_{t_1} - \tilde{X}_{t_2}\right)^2\right]^{\frac{1}{2}} \\
&\leqslant \left(\mathbf{E}\left[h^2(X_{n,s'_1}, \dots, X_{n,s'_v})\right]\right)^{\frac{1}{2}} L\vartheta'_r \quad \text{(c.f. the comp. regarding the autocovariances)} \\
&= \left(\mathbf{E}\left[f^2(Z_{n,s_1}, \dots, Z_{n,s_u})\right]\right)^{\frac{1}{2}} L\vartheta^*_{r-1}
\end{aligned}
$$

Now, consider the case $r' = 0$, i.e. $r = 1$. Then, $s'_v = t'_1$ and $s_u = t'_1 = t_1 - 1$ and

$$
\begin{aligned}
&\left|\mathbf{Cov}\left[h\left(X_{n,s'_1}, X_{n,s'_2}, \dots, X_{n,s'_v}\right), g(X_{n,t'_1}, X_{n,t'_2})\right]\right| \\
&\qquad \leqslant \left(\mathbf{E}\left[h^2(X_{n,s'_1}, \dots, X_n, s'_v)\right]\right)^{\frac{1}{2}} \left(\mathbf{E}[Z^2_{s'_v+1}]\right)^{\frac{1}{2}} \leqslant \left(\mathbf{E}\left[f^2(Z_{n,s_1}, \dots, Z_n, s_u)\right]\right)^{\frac{1}{2}} \kappa
\end{aligned}
$$

This yields the proposition by defining $\vartheta_r = \kappa$ for $r = 1$ and $\vartheta_r = \vartheta^*_r$ for $r \geqslant 2$.

Weak-dependance condition II: Let f be measurable and bounded and $s_1 < \dots < s_u < s_u + r = t_1 \leqslant t_2$. Similar to the first condition we obtain for $r \geqslant 2$:

$$
\begin{aligned}
&\left|\mathbf{Cov}\left[f\left(Z_{n,s_1}, \dots, Z_{n,s_u}\right), Z_{n,t_1} Z_{n,t_2}\right]\right| \\
&\quad = \left|\mathbf{Cov}\left[h\left(X_{n,s'_1}, \dots, X_{n,s'_v}\right), g\left(X_{n,t'_1}, \dots, X_{n,t'_4}\right) - g\left(\tilde{X}_{n,t'_1}, \dots, \tilde{X}_{n,t'_4}\right)\right]\right| \\
&\quad \leqslant 2 \sup_x |h(x)| \mathbf{E}\left[\left|g\left(X_{n,t'_1}, \dots, X_{n,t'_4}\right) - g\left(\tilde{X}_{n,t'_1}, \dots, \tilde{X}_{n,t'_4}\right)\right|\right] \\
&\quad \leqslant 2 \sup_x |h(x)| \mathbf{E}\left[\frac{L}{h\varepsilon}\left|\left(X_{n,t'_1}, \dots, X_{n,t'_4}\right) - \left(\tilde{X}_{n,t'_1}, \dots, \tilde{X}_{n,t'_4}\right)\right|\right] \\
&\quad \leqslant 2 \sup_x |h(x)| L\vartheta^*_r
\end{aligned}
$$

Now, consider the case $r' = 0$, i.e. $r = 1$. Then, $s'_v = t'_1$ and $s_u = t'_1 = t_1 - 1$ and

$$
\begin{aligned}
&\left|\mathbf{Cov}\left[h(X_{n,s'_1}, \dots, X_{n,s'_v}, g(X_{n,s'_v}, X_{n,s'_v+1}), g(X_{n,t'_2-1}, X_{n,t_2}\right]\right| \\
&\qquad = \left|\mathbf{Cov}\left[h(X_{n,s'_1}, \dots, X_{n,s'_v}, Z_{n,s'_v+1} Z_{n,t'_2}\right]\right| \\
&\qquad \leqslant 2 \sup_x |h(x)| \mathbf{E}\left[|Z_{n,s'_v+1} Z_{n,t'_2}|\right] \\
&\qquad = 2 \sup_x |h(x)| \kappa
\end{aligned}
$$

This yields the proposition by defining $\vartheta_r = \kappa$ for $r = 1$ and $\vartheta_r = \vartheta^*_r$ for $r \geqslant 2$. □

We conclude the considerations regarding this estimator with the

Proof of Theorem 5.1.9. As we have seen, we have (neglect terms with higher order of h and ε)

$$
\begin{aligned}
MSE_h &= c_1 \frac{1}{nh\varepsilon} + c_2 h^4 + c_3 \varepsilon^4 + 2c_4 h^2 \varepsilon^2 = c_1 \frac{1}{nh^{\alpha+1}} + c_2 h^4 + c_3 h^{4\alpha} + 2c_4 h^{2\alpha+2} \\
\Longrightarrow \qquad 0 &\stackrel{!}{=} c_1 \frac{\alpha+1}{nh^{\alpha+2}} + 4\left(c_2 h^3 + \alpha c_3 h^{4\alpha-1} + (\alpha+1) c_4 h^{2\alpha+1}\right) \\
\Longleftrightarrow \qquad \frac{c_5}{n} &= c_2 h^{5+\alpha} + c_3 \alpha h^{5\alpha+1} + c_4(\alpha+1) h^{3\alpha+3}
\end{aligned}
$$

with $c_5 = -c_1 \frac{\alpha+1}{4}$. For $\alpha \geqslant 1$ the term $h^{5+\alpha}$ dominates, whereas for $\alpha \leqslant 1$ the term $h^{5\alpha+1}$ dominates. Hence,

$$\frac{c_5}{n} = \begin{cases} h^{5+\alpha}, & \alpha \geqslant 1 \\ \alpha h^{5\alpha+1}, & \alpha \leqslant 1 \end{cases} \quad \Longrightarrow \quad h^* = \begin{cases} \mathcal{O}\left(n^{-\frac{1}{5+\alpha}}\right), & \alpha \geqslant 1 \\ \mathcal{O}\left(n^{-\frac{1}{5\alpha+1}}\right), & \alpha \leqslant 1 \end{cases}$$

and

$$MSE_h^*(y) = c_1 \frac{1}{nh^{\alpha+1}} + c_2 h^4 + c_3 h^{4\alpha} = \begin{cases} (c_1+c_2)n^{-\frac{4}{5+\alpha}}, & \alpha \geqslant 1 \\ (c_1+c_3)n^{-\frac{4\alpha}{5\alpha+1}}, & \alpha \leqslant 1 \end{cases}.$$

Since

$$\begin{aligned} \frac{\partial}{\partial \alpha} n^{-\frac{4}{5+\alpha}} &= n^{-\frac{4}{\alpha+5}} \ln n \frac{4}{(\alpha+5)^2} > 0, \\ \frac{\partial}{\partial \alpha} n^{-\frac{4\alpha}{5\alpha+1}} &= n^{-\frac{4}{\alpha+5}} \ln n \left(-\frac{4(5\alpha+1) - 4\alpha - 5}{(\alpha+5)^2} \right) = -n^{-\frac{4}{\alpha+5}} \ln n \frac{1}{(\alpha+5)^2} < 0 \end{aligned}$$

the *MSE* is minimal for $\alpha = 1$, i.e. $\varepsilon = \mathcal{O}(h)$ and the assertion directly follows. □

5.6.2 Disturbance parameter

Proof of Theorem 5.3.2. Let us first neglect the norming factor. According to Equation (5.1) we can split up the kernel K into three parts by a taylor expansion and obtain for the first term:

$$\begin{aligned} &\mathbf{E}\left[\frac{1}{nh\varepsilon} \sum_{r=1}^{n} \int e^{ity} K\left(\frac{y - b_t X_{t-1} - e_t}{h} \right) \mathrm{d}y W\left(\frac{X_{r-1}}{\varepsilon} \right) \right] \\ &\quad = \frac{1}{\varepsilon} \int\int\int\int K(u) W\left(\frac{x}{\varepsilon}\right) e^{(hu+bx+e)it} f(x)h(b)k(e) \mathrm{dxdbdedu} \\ &\quad = \int e^{ite} k(e) \mathrm{de} \int e^{ituh} K(u) \mathrm{du} \int\int e^{itv\varepsilon b} f(v\varepsilon) h(b) W(v) \mathrm{dbdv} \\ &\quad = \phi_e(t) \left(1 + \frac{1}{2} h^2 t^2 \int u^2 K(u) \mathrm{du} + \mathcal{O}\left(h^3 t^3\right) \right) \left(f(0) + \mathcal{O}\left(\varepsilon^2\right) \right) \end{aligned}$$

The last step can be seen as follows, with $\vartheta \in [0,1]$:

$$\int e^{ituh} K(u) \mathrm{du} = \int \left(1 + ithu + \frac{(ith)^2}{2} u^2 + \frac{(ith)^3}{6} u^3 e^{i\vartheta tuh} \right) K(u) \mathrm{du} = 1 + \mathcal{O}\left(t^3 h^3\right)$$

and for the term $\int\int e^{itv\varepsilon b} f(v\varepsilon) h(b) W(v) \mathrm{dbdv}$ we perform a taylor expansion for the exponential function and one for f. For the variance we obtain by the following argumentation and the argumentation used in the proof of Lemma 5.3.5 the order $\mathcal{O}\left(\frac{1}{nh\varepsilon}\right)$:

$$\frac{1}{h^2\varepsilon^2}\mathbf{E}\left[\left(\left(\int \mathrm{e}^{ity} K\left(\frac{y-b_tX_{t-1}-e_t}{h}\right)\mathrm{dy}W\left(\frac{X_{r-1}}{\varepsilon}\right)\right)^2\right]$$
$$= \frac{1}{h\varepsilon^2}\int\int\int\left(K(u)W\left(\frac{x}{\varepsilon}\right)\mathrm{e}^{(hu+bx+e)it}\mathrm{du}\right)^2 f(x)h(b)k(e)\mathrm{dxdbde}$$
$$\leqslant \frac{1}{h\varepsilon^2}\int\int\int\int K^2(u)W^2\left(\frac{x}{\varepsilon}\right)f(x)h(b)k(e)\mathrm{dudxdbde}$$
$$= \frac{1}{h\varepsilon}\int K^2(u)\mathrm{du}\int W^2(v)\mathrm{dv}\left(f(0)+\mathcal{O}(\varepsilon^2)\right)$$

By the argumentation of Lemmas 5.1.5 and 5.1.6 the other terms that we obtain by the taylor expansion of the kernel (Equation (5.1)) are neglectable. By Lemma 5.1.4 we know that the norming factor is equal to $f(0)+\mathcal{O}\left(\varepsilon^2+\frac{1}{\sqrt{n\varepsilon}}\right)$, so that the Theorem is proved. □

Smooth case: Consistency

Proof of Lemma 5.3.5. We proceed in the same way as in the proof of Lemma 5.1.3:

$$\mathbf{E}\left[\tilde{h}\left(\frac{z}{a}\right)\right] = \int\int\int\frac{|a|}{2\pi}\frac{1}{\delta}\int e^{it(bx+e-z)}\frac{\phi_G(tk)}{\phi_e(t)}\mathrm{dt}V\left(\frac{x-a}{\delta}\right)k(e)\mathrm{de}h(b)f(x)\mathrm{dbdx}$$
$$= \int\int\frac{|a|}{\delta}G(u)\frac{1}{x}V\left(\frac{x-a}{\delta}\right)h\left(\frac{z-uk}{x}\right)f(x)\mathrm{dudx}$$
$$= \int\int|a|G(u)V(v)\frac{1}{\delta v+a}f(\delta v+a)h\left(\frac{z-uk}{\delta v+a}\right)\mathrm{dudv}$$
$$= f(a)h\left(\frac{z}{a}\right)+\mathcal{O}\left(k^2+\delta^2\right)$$

To bound the sum

$$\sum_{j=2}^{\infty}\mathbf{Cov}\left[\frac{1}{2\pi}\int e^{it\frac{b_1X_0+e_1-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(\frac{X_0-a}{\delta}\right),\right.$$
$$\left.\frac{1}{2\pi}\int e^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(\frac{X_{j-1}-a}{\delta}\right)\right]$$

we split it up into the sum of the first $M=M(n)\overset{n\to\infty}{\longrightarrow}\infty$ summands and the residual sum. Exactly in the same way as before, we obtain that the sum of the first M autocovariances is of order $\mathcal{O}\left(Mk^2\delta^2\right)$ and that we have one term of order $\mathcal{O}(\frac{k\delta}{k^{2\beta}})$. To evaluate the residual sum we note that

$$\frac{1}{k^2\delta^2}\frac{1}{4\pi^2}\mathbf{E}\left[\left(\left(\int e^{-it\frac{-e-bx+z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(\frac{X_{j-1}-a}{\delta}\right)\right)\right)^2\right]$$
$$= \frac{1}{\delta^2}\frac{1}{4\pi^2}\int\int\int\left(\frac{1}{k}\int e^{-it\frac{-e-bx+z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}\right)^2V^2\left(\frac{x-a}{\delta}\right)k(e)\mathrm{de}h(b)f(x)\mathrm{dbdx} \quad (5.14)$$
$$= \frac{1}{\delta^2}\frac{1}{4\pi^2}\int\int\int\frac{1}{k}\left(\int e^{-itu}\frac{t^\beta}{k^\beta}\phi_G(t)\mathrm{dt}+\int_{-Mk}^{Mk}\mathrm{e}^{-itu}\left(\frac{1}{\phi_e\left(\frac{t}{k}\right)}-\frac{t^\beta}{k^\beta}\right)\phi_G(t)\mathrm{dt}\right)^2\mathrm{du}$$

$$V^2\left(\frac{x-a}{\delta}\right)\left(k(z-bx)+\mathcal{O}(k)\right)h(b)f(x)\mathrm{dbdx}$$

$$= \frac{1}{\delta^2}\frac{1}{4\pi^2}\iiint \frac{1}{k^{2\beta+1}}\left[\left(\int e^{-itu}t^\beta\phi_G(t)\mathrm{dt}\right)^2 + \left(\int_{-Mk}^{Mk}\mathrm{e}^{-itu}\left(\frac{k^\beta}{\phi_e\left(\frac{t}{k}\right)}-t^\beta\right)\phi_G(t)\mathrm{dt}\right)^2\right.$$

$$\left.+2\left(\int e^{-itu}t^\beta\phi_G(t)\mathrm{dt}\right)\left(\int_{-Mk}^{Mk}\mathrm{e}^{-itu}\left(\frac{k^\beta}{\phi_e\left(\frac{t}{k}\right)}-t^\beta\right)\phi_G(t)\mathrm{dt}\right)\right]\mathrm{du}$$

$$V^2\left(\frac{x-a}{\delta}\right)\left(k(z-bx)+\mathcal{O}(k)\right)h(b)f(x)\mathrm{dbdx}$$

$$= \frac{1}{\delta^2}\frac{1}{4\pi^2}\int\int\int \frac{1}{k^{2\beta+1}}\left[\left(\int e^{-itu}t^\beta\phi_G(t)\mathrm{dt}\right)^2 - \left(\int_{-Mk}^{Mk}\mathrm{e}^{-itu}t^\beta\phi_g(t)\mathrm{dt}\right)^2\right.$$

$$\left.-2\left(\int e^{-itu}t^\beta\phi_G(t)\mathrm{dt}\right)\left(\int_{-Mk}^{Mk}\mathrm{e}^{-itu}t^\beta\phi_g(t)\mathrm{dt}\right)\right]\mathrm{du}+\mathcal{O}(k^\beta)$$

$$V^2\left(\frac{x-a}{\delta}\right)\left(k(z-bx)+\mathcal{O}(k)\right)h(b)f(x)\mathrm{dbdx}$$

$$= \frac{1}{\delta^2}\frac{1}{4\pi^2}\int\int\int \frac{1}{k}\left(\underbrace{\left(\int e^{-itu}\frac{t^\beta}{k^\beta}\phi_G(t)\mathrm{dt}\right)^2\mathrm{du}}_{Parseval\ on\ F(t)=t^\beta\phi_G(t)}+\frac{1}{k^{2\beta}}\mathcal{O}(k)\right)$$

$$V^2\left(\frac{x-a}{\delta}\right)\left(k(z-bx)+\mathcal{O}(k)\right)h(b)f(x)\mathrm{dbdx} \tag{5.15}$$

$$= \int\int \frac{1}{k^{2\beta+1}\delta}V^2(v)k(z-b\delta v+ba)h(b)f(\delta v+a)\mathrm{dvdb}\left(\int t^{2\beta}\phi_G^2(t)\mathrm{dt}+\mathcal{O}(k)\right)$$

$$= \frac{1}{k^{2\beta+1}\delta}f(a)\int V^2(v)\mathrm{dv}\int k(z+ba)h(b)\mathrm{db}\int t^{2\beta}\phi_G^2(t)\mathrm{dt}+\mathcal{O}\left(\frac{k+\delta}{k^{2\beta+1}\delta}\right)$$

$$= \mathcal{O}\left(\frac{1}{k^{2\beta+1}\delta}\right).$$

Further, we conclude that for a Lipschitz continuous function V the function

$$F(b,c,d) = \frac{1}{2\pi}\int \mathrm{e}^{it\left(b-\varphi c-\frac{z}{k}\right)}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(d-\frac{a}{\delta}\right)$$

is Lipschitz continuous as well if $\int |t|^{\beta+1}\,|\phi_G(t)| < \infty$ since it holds with $u_i = b_i - \varphi_i$:

$$\left|\frac{1}{2\pi}\int\left(\mathrm{e}^{itu_1}-\mathrm{e}^{itu_2}\right)\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}\right|$$

$$= \frac{1}{2\pi}\left|\int\left(\cos(tu_1)+i\sin(tu_1)-\cos(tu_2)-i\sin(tu_2)\right)\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}\right|$$

$$= \frac{L}{\pi}\int\underbrace{\left|\sin\left(\frac{t}{2}(u_1-u_2)\right)\right|}_{\leqslant L\frac{t}{2}(u_1-u_2)}\underbrace{\left(\left|\cos\left(\frac{tu_1+tu_2}{2}\right)\right|+\left|\sin\left(\frac{tu_1+tu_2}{2}\right)\right|\right)}_{\leqslant 2}\left|\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\right|\mathrm{dt}$$

$$\leqslant (u_1-u_2)\frac{L}{\pi}\int|t|\left|\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\right|\mathrm{dt}$$

$$= \frac{u_1 - u_2}{k^\beta} \frac{1}{\pi} \int |t|^{\beta+1} |\phi_G(t)| \mathrm{d}t + \mathcal{O}(k).$$

Moreover, a geometrically decaying bound for a term of a squared expectation is given by the following expression, where we denote the Lipschitz constant of F by L_F:

$$\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{2\pi}\int e^{it\frac{b_j X_{j-1}+e_j-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_{j-1}-a}{\delta}\right)\right.\right. \\
&\qquad\left.\left. -\frac{1}{2\pi}\int e^{it\frac{b_j \tilde{X}_{j-1}^{j-1}+e_j-z}{k}} \frac{\tilde{\phi}_G^{j-1}(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{\tilde{X}_{j-1}^{j-1}-a}{\delta}\right)\right)^2\right]^{\frac{1}{2}} \\
&= \mathbf{E}\left[\left(\frac{1}{2\pi}\int e^{it\frac{X_j-\varphi X_{j-1}-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_{j-1}-a}{\delta}\right)\right.\right. \\
&\qquad\left.\left. -\frac{1}{2\pi}\int e^{it\frac{\tilde{X}_j^{j-1}-\varphi \tilde{X}_{j-1}^{j-1}+e_j-z}{k}} \frac{\tilde{\phi}_G^{j-1}(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{\tilde{X}_{j-1}^{j-1}-a}{\delta}\right)\right)^2\right]^{\frac{1}{2}} \\
&= \mathbf{E}\left[\left(F\left(\frac{X_j}{k}, \frac{X_{j-1}}{k}, \frac{X_{j-1}}{\delta}\right) - F\left(\frac{\tilde{X}_j^{j-1}}{k}, \frac{\tilde{X}_{j-1}^{j-1}}{k}, \frac{\tilde{X}_{j-1}^{j-1}}{\delta}\right)\right)^2\right]^{\frac{1}{2}} \\
&\leqslant \frac{L_F}{k^\beta}\mathbf{E}\left[\left(\frac{X_j - \tilde{X}_j^{j-1}}{k}\right)^2 + \left(\frac{X_{j-1} - \tilde{X}_{j-1}^{j-1}}{k}\right)^2 + \left(\frac{X_{j-1} - \tilde{X}_{j-1}^{j-1}}{\delta}\right)^2\right]^{\frac{1}{2}} \\
&= \frac{L_F}{k^\beta}\sqrt{\left(\frac{2}{k^2}+\frac{1}{\delta^2}\right)\frac{\sigma^2}{1-\vartheta}\vartheta^{j-1}} \\
&= \sqrt{\vartheta^{j-1}}\frac{L_F}{k^\beta}\frac{k+\delta}{k\delta}\sqrt{\frac{2\sigma^2}{1-\vartheta}}.
\end{aligned}$$

Using these considerations, we obtain for the residual sum

$$\begin{aligned}
&\sum_{j=M+1}^{\infty} \mathbf{Cov}\left[\frac{1}{2\pi}\int e^{it\frac{b_1 X_0+e_1-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_0-a}{\delta}\right),\right. \\
&\qquad \left(\frac{1}{2\pi}\int e^{it\frac{b_j X_{j-1}+e_j-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_{j-1}-a}{\delta}\right)\right. \\
&\qquad\qquad \left.\left. -\frac{1}{2\pi}\int e^{it\frac{b_j \tilde{X}_{j-1}^{j-1}+e_j-z}{k}} \frac{\tilde{\phi}_G^{j-1}(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{\tilde{X}_{j-1}^{j-1}-a}{\delta}\right)\right)\right] \\
&\leqslant \sum_{j=M+1}^{\infty} \mathbf{E}\left[\left(\frac{1}{2\pi}\int e^{it\frac{b_1 X_0+e_1-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_0-a}{\delta}\right)\right)^2\right]^{\frac{1}{2}} \\
&\qquad \cdot \mathbf{E}\left[\left(\frac{1}{2\pi}\int e^{it\frac{b_j X_{j-1}+e_j-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} \mathrm{d}t V\left(\frac{X_{j-1}-a}{\delta}\right)\right.\right.
\end{aligned}$$

$$-\frac{1}{2\pi}\int e^{it\frac{b_j\tilde{X}_{j-1}^{j-1}+e_j-z}{k}}\frac{\tilde{\phi}_G^{j-1}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{\tilde{X}_{j-1}^{j-1}-a}{\delta}\right)\Bigg)^2\Bigg]^{\frac{1}{2}}$$

$$=\sqrt{\vartheta^M}\frac{k+\delta}{k^{2\beta}\sqrt{k\delta}}C_1L_F\sqrt{\frac{2\sigma^2}{1-\vartheta}}\frac{1}{1-\sqrt{\vartheta}}$$

and can evaluate the estimator's variance:

$$\begin{aligned}
\mathbf{Var}\left[\tilde{h}\left(\frac{z}{a}\right)\right] &= \mathbf{Var}\left[\frac{1}{n\delta}\sum_{j=1}^{n}\frac{|a|}{2\pi}\int e^{it(b_jX_{j-1}+e_j-z)}\frac{\phi_G(tk)}{\phi_e(t)}\mathrm{d}tV\left(\frac{X_{j-1}-a}{\delta}\right)\right]\\
&\leqslant \frac{1}{nk^2\delta^2}\frac{a^2}{4\pi^2}\mathbf{E}\left[\left(\left(\int e^{it\frac{b_1X_0+e_1-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_0-a}{\delta}\right)\right)\right)^2\right]\\
&\quad+\frac{2}{n^2k^2\delta^2}\sum_{j=1}^{n}\underbrace{(n-i+1)}_{\leqslant n}\frac{a^2}{4\pi^2}\mathbf{Cov}\left[\int e^{it\frac{b_1X_0+e_1-z}{k}}\frac{\phi_G(k)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_0-a}{\delta}\right),\right.\\
&\qquad\left.\int e^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{\phi_G(k)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_{j-1}-a}{\delta}\right)\right]\\
&=\mathcal{O}\left(\frac{1}{nk^{2\beta+1}\delta}+\frac{M}{n}+\alpha^M\frac{\delta+k}{nk^{2\beta+2.5}\delta^{2.5}}+\frac{1}{nk^{2\beta+1}\delta}\right)
\end{aligned}$$

□

We conclude this part with the proof of the Lemmas regarding the terms $T(z)$ and $S(z)$.

Proof of Lemma 5.3.7. We note that we assumed $\hat{\varphi}$ to be $\sqrt{n}$-consistent for φ and rewrite the summand with T_1 first:

$$\begin{aligned}
&\frac{1}{2\pi}\int \mathrm{e}^{-itz}\frac{\int \mathrm{e}^{its}T_1(s)\mathrm{d}s}{\phi_e(t)}\mathrm{d}t\\
&=\frac{\varphi-\hat{\varphi}}{k}\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{-itz}\frac{\int \mathrm{e}^{its}G'\left(\frac{s-b_jX_{j-1}-e_j}{k}\right)\mathrm{d}s}{\phi_e(t)}\mathrm{d}t\\
&=\frac{\varphi-\hat{\varphi}}{k}\frac{1}{n\delta}\sum_{j=1}^{n}X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{-itz}\frac{\int \mathrm{e}^{it(b_jX_{j-1}+e_j)}F_{G'}(tk)}{\phi_e(t)}\mathrm{d}t\\
&=\frac{\varphi-\hat{\varphi}}{k}\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\\
&=\mathcal{O}_P\left(\frac{1}{k\sqrt{n}}+\frac{1}{nk^{\beta+1}\sqrt{k\delta}}\right),
\end{aligned}$$

because the asymptotic behavior of the sum is as follows:

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\right]\\
&=\int\int\frac{1}{k\delta}xV\left(\frac{x-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{-it\frac{z-bx}{k}}\int \mathrm{e}^{i\frac{t}{k}e}k(e)\mathrm{d}e\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}th(b)f(x)\mathrm{d}b\mathrm{d}x\\
&=\int\int\frac{1}{k\delta}xV\left(\frac{x-a}{\delta}\right)G'\left(\frac{z-bx}{k}\right)h(b)f(x)\mathrm{d}b\mathrm{d}x=\mathcal{O}(1),
\end{aligned}$$

by taylor expansions and substitutions and further

$$\begin{aligned}
&\left|\mathbf{Var}\left[\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\right]\right| \\
&\quad\leqslant\left|\frac{1}{n}\mathbf{E}\left[\frac{1}{k^2\delta^2}X_0^2V^2\left(\frac{X_{j-1}-a}{\delta}\right)\left(\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\right)^2\right]\right| \\
&\qquad+\left|\frac{2}{nk^2\delta^2}\sum_{j=2}^{n}\mathbf{Cov}\left[X_0V\left(\frac{X_0-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_1X_0+e_1-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t,\right.\right. \\
&\qquad\qquad\left.\left.X_{j-1}V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{it\frac{b_jX_{j-1}+e_j-z}{k}}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\right]\right| \\
&\quad=\mathcal{O}\left(\frac{1}{nk^{2\beta+1}\delta}\right)
\end{aligned}$$

by the same argumentation as in the proof of Lemma 5.3.5 and the facts that

$$\begin{aligned}
F_{G'}(t) &= \int \mathrm{e}^{its}G'(s)\mathrm{d}s = \left[\mathrm{e}^{its}G(s)\right]-\int it\mathrm{e}^{its}G(s)\mathrm{d}s = -it\phi_G(t), \\
F_{G''}(t) &= \int \mathrm{e}^{its}G''(s)\mathrm{d}s = \left[\mathrm{e}^{its}G'(s)\right]-\int it\mathrm{e}^{its}G(s)\mathrm{d}s = -itF_{G'}(t) = -(it)^2\phi_G(t) = t^2\phi_G(t).
\end{aligned}$$

Now, we can rewrite the expression with T_2 and obtain by the same argumentation as before that

$$\begin{aligned}
&\frac{1}{2\pi}\int \mathrm{e}^{-itz}\frac{\int \mathrm{e}^{its}T_2(s)\mathrm{d}s}{\phi_e(t)}\mathrm{d}t \\
&\quad=\left(\frac{\varphi-\hat{\varphi}}{k}\right)^2\frac{1}{nk\delta}\sum_{j=1}^{n}X_{j-1}^2V\left(\frac{X_{j-1}-a}{\delta}\right)\frac{1}{2\pi}\int \mathrm{e}^{\frac{it}{k}\left(b_jX_{j-1}+e_j-z+\vartheta\frac{\varphi-\hat{\varphi}}{k}X_{j-1}\right)}\frac{F_{G'}(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t \\
&\quad=\mathcal{O}_P\left(\frac{1}{k^3n}+\frac{1}{nk^{\beta+2}\sqrt{nk\delta}}\right)
\end{aligned}$$

Combining the results for T_1 and T_2 directly yields the desired result. □

We conclude this section with the

Proof of Lemma 5.3.8. We first note that under the assumptions we made, RCA processes are strong mixing at a geometrically decaying rate. The processes we consider here are special variants of the processes that were considered in Pham (1986) and Feigin & Tweedie (1985). The argumentation used there (for example Theorem 3) yields that the processes considered here are strong mixing in the sense of Rosenblatt (1971) as well. This means that $\sup_{B\in\sigma(X_0),\,F\in\sigma(X_n)}|P(B\cap F)-P(B)P(F)| =: d(n)$ and $d(n)=D\cdot d^n$ with $d<1$. If we denote by $f_{r,j}$ the density of the joint distribution of X_r and X_j and by f the distributions of X_r and X_j that are actually identical due to the strong stationarity of the process $(X_t)_{t\in\mathbb{Z}}$, we have especially $f_{r,j}(a,0)=f(a)f(0)+C\lambda^{|r-j|}$ with $|\lambda|<1$. We recall that we assumed $\int|\phi_G(s)|s^3\mathrm{ds}$ to be finite. With the results of Theorem 5.3.2 we have:

$$
\begin{aligned}
S_1(z) &= \frac{|a|}{2\pi}\int\frac{1}{nk\delta}\sum_{r=1}^{n}\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\\
&= \frac{|a|}{2\pi}\int\frac{1}{nk\delta}\sum_{r=1}^{n}\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\\
&\qquad\frac{\frac{1}{A_2}\frac{1}{nh\varepsilon}\sum_{j=1}^{n}\int\mathrm{e}^{i\frac{s}{k}y}K\left(\frac{y-b_tX_{t-1}-e_t}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)\mathrm{dy}-\phi_e\left(\frac{s}{k}\right)}{\phi_e\left(\frac{s}{k}\right)\left(1+\frac{h^2s^2}{k^2}J_1+\mathcal{O}_P\left(\frac{s^3h^3}{k^3}+\varepsilon^2\right)\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\\
&= \frac{|a|}{2\pi A_2}\iint\frac{1}{n^2hk\varepsilon\delta}\sum_{r=1}^{n}\sum_{j=1}^{n}\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)^2}\left(1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)\\
&\qquad\left(\mathrm{e}^{i\frac{s}{k}y}K\left(\frac{y-b_tX_{t-1}-e_t}{h}\right)W\left(\frac{X_{t-1}}{\varepsilon}\right)-\phi_e\left(\frac{s}{k}\right)\right)\mathrm{dsdy}V\left(\frac{X_{r-1}-a}{\delta}\right)
\end{aligned}
$$

where we have set $J_1=\frac{1}{8}\int u^2K(u)\mathrm{du}$. We can split up this term into two summands and several terms of higher order. For the first one's expectation we obtain:

$$
\begin{aligned}
&\frac{1}{n^2k\varepsilon\delta2\pi}\sum_{r=1}^{n}\sum_{j=1}^{n}\int\int\int\int\int\int\int\int\mathrm{e}^{i\frac{s}{k}(b_rx_{r-1}+e_r-z+b_jxj-1+e_j+hu)}K(u)W\left(\frac{x_{j-1}}{\varepsilon}\right)V\left(\frac{x_{r-1}-a}{\delta}\right)\\
&\qquad\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)^2}k(e_r)k(e_j)h(b_r)h(b_j)f_{r,j}(x_{r-1},x_{j-1})\mathrm{db}_r\mathrm{db}_j\mathrm{de}_r\mathrm{de}_j\mathrm{dx}_{r-1}\mathrm{dx}_{j-1}\mathrm{duds}\\
&= \frac{1}{n^2k\varepsilon\delta2\pi}\sum_{r=1}^{n}\sum_{j=1}^{n}\int\phi_{e_r}\left(\frac{s}{k}\right)\phi_{e_j}\left(\frac{s}{k}\right)\int\mathrm{e}^{i\frac{s}{k}hu}K(u)\mathrm{du}\int\int\int\int\mathrm{e}^{-is\frac{z-b_rx_{r-1}-b_jx_{j-1}}{k}}\\
&\qquad\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)^2}W\left(\frac{x_{j-1}}{\varepsilon}\right)V\left(\frac{x_{r-1}-a}{\delta}\right)h(b_r)h(b_j)f_{r,j}(x_{r-1},x_{j-1})\mathrm{db}_r\mathrm{db}_j\mathrm{dx}_{r-1}\mathrm{dx}_{j-1}\mathrm{ds}\\
&= \frac{1}{n^2\varepsilon\delta2\pi}\sum_{r=1}^{n}\sum_{j=1}^{n}\int\int\int\int\int\mathrm{e}^{-isu}\phi_G(s)\left(1+\frac{h^2s^2}{k^2}J_1+\mathcal{O}\left(\frac{s^3h^3}{k^3}+\varepsilon^2\right)\right)\mathrm{ds}\frac{1}{x_{r-1}}\\
&\qquad W\left(\frac{x_{j-1}}{\varepsilon}\right)V\left(\frac{x_{r-1}-a}{\delta}\right)h\left(\frac{z-ku-b_jx_{j-1}}{x_{r-1}}\right)h(b_j)f_{r,j}(x_{r-1},x_{j-1})\mathrm{dudb}_j\mathrm{dx}_{r-1}\mathrm{dx}_{j-1}\\
&= \frac{1}{n^2}\sum_{r=1}^{n}\sum_{j=1}^{n}\int\int\int\int\left(G(u)+\frac{h^2}{k^22\pi}J_1\int\mathrm{e}^{-isu}s^2\phi_G(s)\mathrm{ds}+\mathcal{O}\left(\frac{h^3}{k^3}+\varepsilon^2\right)\right)\frac{1}{\delta v_r+a}\\
&\qquad W(w_j)V(v_r)h\left(\frac{z-ku-\varepsilon b_jw_j}{\delta v_r+a}\right)h(b_j)f_{r,j}(\delta v_r+a,\varepsilon w_j)\mathrm{dudb}_j\mathrm{dv}_r\mathrm{dw}_j\\
&= \frac{1}{n^2}\sum_{r=1}^{n}\sum_{j=1}^{n}h\left(\frac{z}{a}\right)f_{r,j}(a,0)+\frac{h^2}{k^2}J_2+\mathcal{O}\left(k^2+\delta^2+\varepsilon^2+\frac{h^3}{k^3}\right)\\
&= h\left(\frac{z}{a}\right)\frac{1}{n^2}\sum_{r=1}^{n}\left(\sum_{\substack{j=1\\j\neq r}}^{n}(f(a)f(0)+CR_{r,j})-f(a)f(0)\right)+\frac{h^2}{k^2}J_2+\mathcal{O}\left(k^2+\delta^2+\varepsilon^2+\frac{h^3}{k^3}\right)\\
&= h\left(\frac{z}{a}\right)f(a)f(0)+\frac{h^2}{k^2}J_2+\frac{2C}{n^2}\sum_{r=1}^{n}\sum_{j=1}^{r-1}R_{r,j}-\frac{1}{n}f(a)f(0)+\mathcal{O}\left(k^2+\delta^2+\varepsilon^2+\frac{h^3}{k^3}\right)
\end{aligned}
$$

$$= h\left(\frac{z}{a}\right) f(a)f(0) + \frac{h^2}{k^2} J_2 + \mathcal{O}\left(\frac{1}{n} + k^2 + \delta^2 + \varepsilon^2 + \frac{h^3}{k^3}\right)$$

because $|R_{r,j}| \leqslant \lambda^{|r-j|}$ and hence

$$\frac{1}{n^2}\sum_{r=1}^{n}\sum_{j=1}^{r-1} R_{r,j} \leqslant \frac{1}{n^2}\sum_{r=1}^{n}\sum_{j=1}^{r-1} \lambda^{r-j} = \frac{1}{n^2}\frac{1-\lambda^n - n}{1-\lambda^{-1}} = \mathcal{O}\left(\frac{1}{n}\right).$$

Similar computations to the proof of Theorem 5.3.5 yield that the expectation of the second summand is equal to $f(0)f(a)h\left(\frac{z}{a}\right) + \mathcal{O}\left(h^2 + k^2 + \varepsilon^2 + \delta^2\right)$, so that we obtain

$$\mathbf{E}\left[S_1(z)\right] = \frac{h^2}{k^2} J_2 + \mathcal{O}\left(\frac{1}{n} + k^2 + \delta^2 + \varepsilon^2 + \frac{h^3}{k^3}\right)\left(1 + \mathcal{O}\left(\frac{h^2}{k^2} + \varepsilon^2 + \frac{1}{\sqrt{nh\varepsilon}}\right)\right).$$

Following the previous computations it can be seen that under Conditions 5.3.4 all integral terms are finite. To determine the variance we note that

$$\begin{aligned}
&S_1(z)\\
&= \frac{|a|}{2\pi}\int \frac{1}{nk\delta}\sum_{r=1}^{n} \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right) - \phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} \mathrm{d}s V\left(\frac{X_{r-1}-a}{\delta}\right)\\
&= \frac{|a|}{2\pi}\int \frac{1}{nk\delta}\sum_{r=1}^{n} \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e^2\left(\frac{s}{k}\right)} \frac{\mathcal{O}_P\left(\frac{s^2h^2}{k^2}\right)}{1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2} + \varepsilon^2 + \frac{1}{\sqrt{nh\varepsilon}}\right)} \mathrm{d}s V\left(\frac{X_{r-1}-a}{\delta}\right)\\
&= \frac{|a|}{2\pi}\int \frac{h^2}{nk^3\delta}\sum_{r=1}^{n} \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} s^2 \frac{\phi_G(s)}{\phi_e^2\left(\frac{s}{k}\right)} \mathcal{O}_P\left(1 + \frac{s^2h^2}{k^2} + \varepsilon^2 + \frac{1}{\sqrt{nh\varepsilon}}\right) \mathrm{d}s V\left(\frac{X_{r-1}-a}{\delta}\right)
\end{aligned}$$

so that again by Parseval's identity

$$\begin{aligned}
\mathbf{E}\left[S_1^2(z)\right] &\leqslant \frac{a^2}{4\pi^2}\int\int \frac{h^4}{nk^5\delta^2}\int\left(\int \mathrm{e}^{-isu}\frac{s^{2+2\beta}}{k^{2\beta}}\phi_G(s)\mathcal{O}_P\left(1+\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\mathrm{d}s\right)^2 \mathrm{d}u\\
&\quad V^2\left(\frac{x-a}{\delta}\right) k(z-bx)h(b)f(x)\mathrm{d}b\mathrm{d}x + \mathcal{O}(k)\\
&= \frac{a^2}{4\pi^2}\int\int \frac{h^4}{nk^{4\beta+5}\delta}\int s^{4+4\beta}\phi_G^2(s)\mathcal{O}_P\left(1+\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\mathrm{d}s\\
&\quad V^2(v)\, k(z-ba-b\delta v)h(b)f(x)\mathrm{d}v\mathrm{d}x + \mathcal{O}(k)\\
&= \mathcal{O}\left(\frac{h^4}{nk^{4\beta+5}\delta}\right)
\end{aligned}$$

By a similar argumentation it can be seen that the term $S_2(z)$ asymptotically is of smaller order than $S_1(z)$. Hence, the assertion follows. □

Smooth case: Asymptotic distribution

Even though we do note state the proof for Theorem 5.3.10 here, we state the following result, that cannot only be used for this proof, but also for several other proofs that will follow.

Lemma 5.6.3. *Let the random variables $Y_{n,j}(z)$ and $U_{n,r}(z)$, $j, r = 1, \dots, n$, $n = 1, 2, \dots$ with $\boldsymbol{E}[Y_{n,j}(z)] = \boldsymbol{E}[U_{n,r}] = 0$ be given by*

$$
\begin{aligned}
Y_{n,j}(z) = \frac{k^\beta}{\sqrt{k\delta}} \Bigg\{ \frac{|a|}{2\pi} \int \mathrm{e}^{it\frac{b_j X_{j-1}+e_j-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} dt V\left(\frac{X_{j-1}-a}{\delta}\right) \\
- \boldsymbol{E}\left[\frac{|a|}{2\pi} \int \mathrm{e}^{it\frac{b_j X_{j-1}+e_j-z}{k}} \frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)} dt V\left(\frac{X_{j-1}-a}{\delta}\right) \right] \Bigg\} \\
U_{n,r}(z) = \frac{k^{2\beta+2}}{h^2\sqrt{k\delta}} \Bigg\{ \frac{|a|}{2\pi} \int \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} ds V\left(\frac{X_{r-1}-a}{\delta}\right) \\
- \boldsymbol{E}\left[\frac{|a|}{2\pi} \int \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} ds V\left(\frac{X_{r-1}-a}{\delta}\right) \right] \Bigg\}.
\end{aligned}
$$

Then, it holds

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{j=1}^n Y_{n,j}(z) &\overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \tilde{\sigma}^2) \quad i.D. \\
\frac{1}{\sqrt{n}} \sum_{r=1}^n U_{n,r}(z) &\overset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \sigma_S^2\right) \quad i.\,D.
\end{aligned}
$$

where $\tilde{\sigma}^2$ and σ_S^2 are given in Theorem 5.3.10 and Lemma 5.3.11, respectively.

Proof. Both the results can be shown similarly to Theorem 5.6.2. □

Proof of Theorem 5.3.11. With $U_{n,r}(z)$ as given in Lemma 5.6.3 we obtain

$$
\begin{aligned}
&\frac{k^{2\beta+2}}{h^2}\sqrt{nk\delta}\left(S(z) - \frac{h^2}{k^2}J_2\right) \\
&= \frac{k^{2\beta+2}}{h^2}\sqrt{nk\delta}\frac{|a|}{2\pi} \int \frac{1}{nk\delta} \sum_{r=1}^n \Bigg\{ \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} \mathrm{d}sV\left(\frac{X_{r-1}-a}{\delta}\right) \\
&\quad -\mathbf{E}\left[\int \mathrm{e}^{i\frac{s}{k}(-z+b_r X_{r-1}+e_r)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} \mathrm{d}sV\left(\frac{X_{r-1}-a}{\delta}\right) \right] \Bigg\} + \frac{k^{2\beta+2}}{h^2}\sqrt{nk\delta} \\
&\quad \cdot \left(\mathbf{E}\left[\frac{|a|}{2k\delta\pi} \int \mathrm{e}^{i\frac{s}{k}(-z+b_1 X_0+e_1)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} \mathrm{d}sV\left(\frac{X_0-a}{\delta}\right) \right] - \frac{h^2}{k^2}J_2 \right) \\
&= \frac{1}{\sqrt{n}} \sum_{r=1}^n U_{n,r}(z) + \frac{k^{2\beta+2}}{h^2}\sqrt{nk\delta} \\
&\quad \cdot \left(\mathbf{E}\left[\frac{|a|}{2k\delta\pi} \int \mathrm{e}^{i\frac{s}{k}(-z+b_1 X_0+e_1)} \frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)} \frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)} \mathrm{d}sV\left(\frac{X_0-a}{\delta}\right) \right] - \frac{h^2}{k^2}J_2 \right).
\end{aligned}
$$

By Theorem 5.6.3 we obtain that the first summand converges to a normally distributed random variable and by the proof of Lemma 5.3.8 that the second summand is of order

$$
\frac{k^{2\beta+2}}{h^2}\sqrt{nk\delta}\,\mathcal{O}\left(\frac{1}{n} + k^2 + \delta^2 + \varepsilon^2 + \frac{h^3}{k^3}\right)\left(1 + \mathcal{O}_P\left(\frac{h^2}{k^2} + \varepsilon^2 + \frac{1}{\sqrt{nh\varepsilon}}\right)\right)
$$

what yields the proposition. □

Proof of Theorem 5.3.13. We recall the decomposition of the estimator from Equation (5.7), the definition of $Y_{n,j}$ and $U_{n,r}$ from Lemma 5.6.3, and the asymptotic behavior of $T(z)$ from Lemma 5.3.7 and note further that A_3 is $\sqrt{n\delta}$-consistent for $f(a)$. Then,

$$\begin{aligned}
&k^\beta\sqrt{nk\delta}\left(\hat{h}_n\left(\frac{z}{a}\right)-h\left(\frac{z}{a}\right)-\frac{h^2}{k_2}J_2\right)\\
&\quad=k^\beta\sqrt{nk\delta}\left(\frac{\tilde{h}_n\left(\frac{z}{a}\right)}{A_3}-h\left(\frac{z}{a}\right)+T(z)+\frac{S(z)}{A_3}-\frac{h^2}{k_2}J_2\right)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{j=1}^n Y_{n,j}+\frac{1}{A_3}\frac{1}{\sqrt{n}}\frac{k^{\beta+2}}{h^2}\sum_{r=1}^n U_{n,r}\\
&\qquad\qquad+k^\beta\sqrt{nk\delta}T(z)+k^\beta\sqrt{nk\delta}\left(\frac{S-\frac{h^2}{k^2}J_2}{A_3}+\frac{h^2}{k^2}J_2\frac{f(a)-A_3}{A_3f(a)}\right)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{r=1}^n\left(Y_{n,r}+\frac{C}{A_3}U_{n,r}\right)+k^\beta\sqrt{nk\delta}\,\frac{S-\frac{h^2}{k^2}J_2}{A_3}+\mathcal{O}_P\left(k^{\beta-1}\sqrt{k\delta}+\frac{1}{k\sqrt{n}}+k^\beta\sqrt{k}\right).
\end{aligned}$$

The second summand is evaluated in the proofs of Theorem 5.3.11 and converges to the Bias B_M in probability. The central limit theorem for weak dependent random variables (Neumann & Paparoditis (2008), Th. 6.1) yields that the first summand converges to a normally distributed random variable with zero mean and variance σ_M^2 as given in the Theorem. The proof is analogous to the one of Theorem 5.6.2, for example, because of what we only evaluate the variance here. With Lemma 5.6.3 we obtain that

$$\begin{aligned}
&\mathbf{E}\left[(Y_{n,r}(z)+U_{n,r}(z))^2\right]\\
&\quad=\mathbf{E}\left[Y_{n,r}^2(z)+C^2U_{n,r}^2(z)+2CY_{n,r}(z)U_{n,r}(z)\right]\\
&\quad=f(a)a^2\int V^2(v)\mathrm{dv}\int k(z+ba)h(b)\mathrm{db}\int t^{2\beta}\phi_G^2(t)\mathrm{dt}\\
&\qquad+C^2\frac{a^2}{4\pi^2}\int s^{4+4\beta}\phi_G^2(s)\mathrm{ds}\int V^2(v)\,\mathrm{dv}\int k(z-ba)h(b)f(x)\mathrm{db}\\
&\qquad+2C(-1)^\beta\frac{|a|}{4\pi^2}h\left(\frac{z}{a}\right)f(a)\int\int\phi_e\left(\frac{w}{k}\right)\mathrm{e}^{iwu}\mathrm{dudw}\int V^2(v)\,\mathrm{dv}\int s^{2\beta+2}\phi_G^2(s)\mathrm{ds}+o(1)
\end{aligned}$$

since the joint expectation of $Y_{n,r}(z)$ and $U_{n,r}(z)$ is given by

$$\begin{aligned}
&\mathbf{E}\left[Y_{n,r}(z)U_{n,r}(z)\right]\\
&\quad=\frac{k^{3\beta+2}}{h^2k\delta}\left(\mathbf{E}\left[\frac{|a|}{2\pi}\int\mathrm{e}^{it\frac{b_rX_{r-1}+e_r-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(\frac{X_{r-1}-a}{\delta}\right)\right.\right.\\
&\qquad\qquad\left.\frac{|a|}{2\pi}\int\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\right]\\
&\qquad+\mathbf{E}\left[\frac{|a|}{2\pi}\int\mathrm{e}^{it\frac{b_rX_{r-1}+e_r-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}V\left(\frac{X_{r-1}-a}{\delta}\right)\right]\\
&\qquad\qquad\left.\mathbf{E}\left[\frac{|a|}{2\pi}\int\mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\frac{\hat{\phi}_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat{\phi}_e\left(\frac{s}{k}\right)}\mathrm{ds}V\left(\frac{X_{r-1}-a}{\delta}\right)\right]\right)
\end{aligned}$$

$$= \frac{k^{3\beta+1}}{h^2\delta}(-1)^\beta \frac{|a|}{4\pi^2}\frac{h^2\delta}{k^{3\beta+1}} h\left(\frac{z}{a}\right) f(a) \int\int \phi_e\left(\frac{w}{k}\right) \mathrm{e}^{iwu}\mathrm{dudw} \int V^2(v)\,\mathrm{dv}$$
$$\int s^{2\beta+2}\phi_G^2(s)\left(1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)\mathrm{ds} + \mathcal{O}(k+\delta)$$
$$+ \frac{k^{3\beta+1}}{h^2\delta}\left(h\left(\frac{z}{a}\right)f(a) + \mathcal{O}\left(k^2+\delta^2\right)\right)\mathcal{O}\left(\frac{1}{n}+k^2+\delta^2+\varepsilon^2+\frac{h^3}{k^3}\right)$$
$$\left(1+\mathcal{O}_P\left(\frac{h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)$$
$$= (-1)^\beta \frac{|a|}{4\pi^2} h\left(\frac{z}{a}\right) f(a)\int\int \phi_e\left(\frac{w}{k}\right)\mathrm{e}^{iwu}\mathrm{dudw}\int V^2(v)\,\mathrm{dv}\int s^{2\beta+2}\phi_G^2(s)\mathrm{ds}$$
$$+ \mathcal{O}_P\left(k+\delta+\frac{h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)$$

because

$$\mathbf{E}\left[\frac{a^2}{4\pi^2}\int \mathrm{e}^{it\frac{b_rX_{r-1}+e_r-z}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{dt}\int \mathrm{e}^{i\frac{s}{k}(-z+b_rX_{r-1}+e_r)}\frac{\phi_G(s)}{\phi_e\left(\frac{s}{k}\right)}\frac{\hat\phi_e\left(\frac{s}{k}\right)-\phi_e\left(\frac{s}{k}\right)}{\hat\phi_e\left(\frac{s}{k}\right)}\mathrm{ds}V^2\left(\frac{X_{r-1}-a}{\delta}\right)\right]$$
$$= \frac{a^2}{4\pi^2}\int\int\int\int\int \mathrm{e}^{i\frac{t+s}{k}(bx+e-z)}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\frac{\phi_G(s)}{\phi_e^2\left(\frac{s}{k}\right)}\frac{\mathcal{O}_P\left(\frac{s^2h^2}{k^2}\right)}{1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)}V^2\left(\frac{x-a}{\delta}\right)$$
$$k(e)h(b)f(x)\mathrm{dsdtdbdedx}$$
$$= \frac{a^2}{4\pi^2}\frac{h^2\delta}{k^{3\beta+1}}\int\int\int\int \phi_e\left(\frac{t+s}{k}\right)\mathrm{e}^{i(t+s)u}t^\beta s^{\beta+2}\phi_G(t)\phi_G(s)\left(1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)$$
$$\frac{1}{\delta v+a}V^2(v)\,h\left(\frac{z-ku}{\delta v+a}\right)f(\delta v+a)\mathrm{dsdtdudv}+\mathcal{O}(k)$$
$$= \frac{|a|}{4\pi^2}\frac{h^2\delta}{k^{3\beta}}\int\int\int\int \phi_e(w)\,\mathrm{e}^{iwku}(-s)^\beta s^{\beta+2}\phi_G(-s)\phi_G(s)$$
$$\left(1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)V^2(v)\,h\left(\frac{z}{a}\right)f(a)\mathrm{dsdwdudv}+\mathcal{O}(k+\delta)$$
$$= (-1)^\beta\frac{|a|}{4\pi^2}\frac{h^2\delta}{k^{3\beta+1}}\int\int\phi_e\left(\frac{w}{k}\right)\mathrm{e}^{iwu}\mathrm{dudw}\int s^{2\beta+2}\phi_G^2(s)\left(1+\mathcal{O}_P\left(\frac{s^2h^2}{k^2}+\varepsilon^2+\frac{1}{\sqrt{nh\varepsilon}}\right)\right)\mathrm{ds}$$
$$h\left(\frac{z}{a}\right)f(a)\int V^2(v)\,\mathrm{dv}+\mathcal{O}(k+\delta).$$

This concludes the proof. □

Proof of Theorem 5.3.15. The results concerning $\tilde{h}(\cdot)$ can be obtained exactly in the same way as the results of Theorem 5.1.9. Hence, we only consider the *MSE* between $\hat{h}(\cdot)$ and $h(\cdot)$. Asymptotically, the Bias B_h is minimized for $h^2 = k^{\beta+2}$ and is then equal to B_M. Then, we have

$$MSE_M(z) = c_1\frac{1}{nk^{2\beta+1}\delta}+c_2k^4+c_3\delta^4+c_4k^2\delta^2+c_5\frac{h^6}{k^6}$$
$$= c_1\frac{1}{nk^{2\beta+1+\alpha}}+c_2k^4+c_3k^{4\alpha}+c_4k^{2+2\alpha}+c_5k^{3\beta}$$

$$\Rightarrow \qquad 0 \overset{!}{=} c_1 \frac{2\beta+\alpha+1}{nk^{2\beta+\alpha+2}} + 4c_2k^3 + 4\alpha c_3 k^{4\alpha-1} + 2(\alpha+1)c_4 k^{2\alpha+1} + 3\beta c_5 k^{3\beta-1}$$

$$\Leftrightarrow \qquad \frac{1}{n} = \tilde{c}_1 k^{2\beta+\alpha+5} + \tilde{c}_2 k^{2\beta+5\alpha+1} + \tilde{c}_3 k^{2\beta+3\alpha+3} + \tilde{c}_4 k^{5\beta+\alpha+1}$$

By the same argumentation as in the proof of Theorem 5.1.9, α should be chosen to be equal to one and this term simplifies to

$$\frac{1}{n} = c_1^* k^{2\beta+6} + c_2^* k^{5\beta+2}.$$

For β smaller or larger than $\frac{4}{3}$ either the first oder the second term asymptotically dominates. Hence, the optimal bandwidth is given by $k^* = n^{-\frac{1}{2\beta+6}}$ for $\beta \leqslant \frac{4}{3}$ and by $k^* = n^{-\frac{1}{5\beta+2}}$ for $\beta > \frac{4}{3}$. $\square$

5.6.3 Multi-point estimator

In Section 5.4 we cited one Lemma that we now state.

Lemma 5.6.4 (The expectation of the squared kernels)**.** *For arbitrary* $0 \leqslant \beta, P, N^* < \infty$, $0 < c_1, c_2 < \infty$ *and the parameters chosen to be* $k = \mathcal{O}\left(n^{-1/r}\right)$, $\delta = \mathcal{O}\left(n^{-1/s}\right)$, $N = \sqrt[n]{P^n + Q^n}$, $Q = \frac{N^*}{\delta}$, *and* $C(n) = 2c\left(1 - \frac{N^*}{N\delta}\right)$ *holds*

$$\frac{1}{k^2\delta^2}\frac{1}{4\pi^2}\boldsymbol{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}|a_i|\int e^{-it\frac{-e-bx+z_i}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}dtV\left(\frac{X_{j-1}-a}{\delta}\right)\right)^2\right]$$
$$= \frac{1}{N^2k^{2\beta+1}\delta}\sum_{i=1}^{N} f(a_i)a_i^2 \int V^2(v)\,dv \int k(z+ba)h(b)\,db \int t^{2\beta}\phi_G^2(t)\,dt + o(1)$$

Proof. Analogous to the proof of Lemma 5.3.5 we obtain

$$\frac{1}{k^2\delta^2}\frac{1}{4\pi^2}\mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}|a_i|\int e^{-it\frac{-e-bx+z_i}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}tV\left(\frac{X_{j-1}-a}{\delta}\right)\right)^2\right]$$
$$= \frac{1}{k^2\delta^2}\frac{1}{4\pi^2}\int\int\int\frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}|a_{i_2}||a_{i_1}|\int \mathrm{e}^{-it\frac{-e-bx+z_{i_1}}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t\int \mathrm{e}^{-it\frac{-e-bx+z_{i_2}}{k}}\frac{\phi_G(t)}{\phi_e\left(\frac{t}{k}\right)}\mathrm{d}t$$
$$V\left(\frac{x-a_{i_1}}{\delta}\right)V\left(\frac{x-a_{i_2}}{\delta}\right)k(e)\mathrm{d}eh(b)f(x)\mathrm{d}b\mathrm{d}x$$
$$= \frac{1}{k\delta}\frac{1}{4\pi^2}\frac{2}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{i_1-1}\Bigg(|a_{i_1}||a_{i_2}|\int\int\int\int\int \mathrm{e}^{-it_1u_i}\phi_G(t_1)\mathrm{e}^{-it_2\left(u_{i_1}+\frac{z_{i_2}-z_{i_1}}{k}\right)}\frac{t_1^\beta}{k^\beta}\phi_G(t_2)\frac{t_2^\beta}{k^\beta}\mathrm{d}t_1\mathrm{d}t_2$$
$$V(v)\,V\left(v+\frac{a_{i_1}-a_{i_2}}{\delta}\right)f(\delta v+a_{i_1})\mathrm{d}u_{i_1}\mathrm{d}v_{i_1}k(z-u_{i_1}k-bx)h(b)\mathrm{d}b\Bigg)$$
$$+\frac{1}{N^2k^{2\beta+1}\delta}\sum_{i=1}^{N}f(a_i)a_i^2\int V^2(v_{i_1})\mathrm{d}v_{i_1}\int k(z+ba)h(b)\mathrm{d}b\int t^{2\beta}\phi_G^2(t)\mathrm{d}t+\mathcal{O}\left(\frac{k+\delta}{Nk^{2\beta+1}\delta}\right)$$
$$= \frac{1}{k\delta}\frac{1}{4\pi^2}\frac{2}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{i_1-1}|a_{i_1}||a_{i_2}|\int\int\int \mathrm{e}^{-i(t_1+t_2)u_{i_1}}\mathrm{d}u_{i_1}\phi_G(t_1)\frac{t_1^\beta}{k^\beta}\phi_G(t_2)\frac{t_2^\beta}{k^\beta}\mathrm{e}^{-it_2\frac{z_{i_2}-z_{i_1}}{k}}\mathrm{d}t_1\mathrm{d}t_2$$

$$
\begin{aligned}
&\int V\left(v_{i_1}\right) V\left(v_{i_1}+\frac{a_{i_1}-a_{i_2}}{\delta}\right) \mathrm{dv} f(a_{i_1}) \int k(z-u_{i_1}k-bx)h(b)\mathrm{db}\,(1+\mathcal{O}(\delta)) \\
&\quad+\frac{1}{N^2k^{2\beta+1}\delta}\sum_{i=1}^{N} f(a_i)a_i^2\int V^2(v_{i_1})\mathrm{dv}_{i_1}\int k(z+ba)h(b)\mathrm{db}\int t^{2\beta}\phi_G^2(t)\mathrm{dt}+\mathcal{O}\left(\frac{k+\delta}{Nk^{2\beta+1}\delta}\right) \\
&\leqslant \frac{1}{k^{2\beta+1}\delta}\frac{1}{4\pi^2}\frac{2}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{i_1-1}|a_{i_1}||a_{i_2}|\int V\left(v_{i_1}\right) V\left(v_{i_1}+\frac{a_{i_1}-a_{i_2}}{\delta}\right)\mathrm{dv}_{i_1} f(a_{i_1})\,(1+\mathcal{O}(k+\delta)) \\
&\iint\left(\frac{1}{|t_1+t_2|}\int k(z-bx)h(b)\mathrm{db}\mathbb{1}_{\{t_1\neq -t_2\}}+\int h(b)\mathrm{db}\mathbb{1}_{\{t_1=t_2\}}\right)\phi_G(t_1)t_1^{\beta}\phi_G(t_2)t_2^{\beta}\mathrm{e}^{-it_2\frac{z_{i_2}-z_{i_1}}{k}}\mathrm{dt}_1\mathrm{dt}_2 \\
&\quad+\frac{1}{N^2k^{2\beta+1}\delta}\sum_{i=1}^{N} f(a_i)a_i^2\int V^2(v_{i_1})\mathrm{dv}_{i_1}\int k(z+ba)h(b)\mathrm{db}\int t^{2\beta}\phi_G^2(t)\mathrm{dt}+\mathcal{O}\left(\frac{k+\delta}{Nk^{2\beta+1}\delta}\right) \\
&\leqslant \frac{K\,C(n)}{k^{2\beta+1}\delta}\,(1+\mathcal{O}(k+\delta)) \\
&\quad+\frac{1}{N^2k^{2\beta+1}\delta}\sum_{i=1}^{N} f(a_i)a_i^2\int V^2(v_{i_1})\mathrm{dv}_{i_1}\int k(z+ba)h(b)\mathrm{db}\int t^{2\beta}\phi_G^2(t)\mathrm{dt}+\mathcal{O}\left(\frac{k+\delta}{Nk^{2\beta+1}\delta}\right)
\end{aligned}
$$

where we have used that the kernel V has support $[-1,1]$ and have set

$$
\begin{aligned}
C(n) &= \int V(v)V\left(v+\frac{a_{i_1}-a_{i_2}}{\delta}\right)\mathrm{dv} = \int_{-1}^{1-\frac{a_{i_1}-a_{i_2}}{\delta}} V(v)V\left(v+\frac{a_{i_1}-a_{i_2}}{\delta}\right)\mathrm{dv} \\
&\leqslant \int_{-1}^{1-\frac{a_{i_1}-a_{i_2}}{\delta}} c\,\mathrm{dv} = c\left(1-\frac{a_{i_1}-a_{i_2}}{\delta}+1\right) \leqslant c\left(2-\frac{2N^*}{N\delta}\right)
\end{aligned}
$$

by construction of the points. It further holds

$$
\begin{aligned}
\frac{1}{k^{2\beta+1}\delta}\left(1-\frac{N^*}{N\delta}\right) = \frac{\sqrt[n]{P^n+Q^n}\,\delta-N^*}{\sqrt[n]{P^n+Q^n}\,\delta^2k^{2\beta+1}} &= \frac{\left(P^n\delta^n+N^{*n}\right)^{1/n}-N^*}{\left(P^n\delta^n+N^{*n}\right)^{1/n}\delta k^{2\beta+1}} \\
&= \frac{\left(\frac{P^n}{c_1^n n^{n/s}}+N^{*n}\right)^{1/n}-N^*}{\left(\frac{P^n}{c_1^n n^{n/s}}+N^{*n}\right)^{1/n}} c_2^{2\beta+1} n^{\frac{2\beta+1}{r}+\frac{1}{s}},
\end{aligned}
$$

where denominator converges to some number κ with $0<\kappa<\infty$, and the numerator can be written as

$$
\begin{aligned}
c_2^{2\beta+1}n^{\frac{2\beta+1}{r}+\frac{1}{s}}N^*\left(\left(\frac{P^n}{N^{*n}c_1^n n^{n/s}}+1\right)^{1/n}-1\right) &\leqslant c_2^{2\beta+1}n^{\frac{2\beta+1}{r}+\frac{1}{s}}N^*\,\frac{P^n}{N^{*n}c_1^n n^{n/s}} \\
&= c_2^{2\beta+1}\frac{R^n}{n^{\frac{n-1}{s}-\frac{2\beta+1}{r}}} = o(1),
\end{aligned}
$$

so that the assertion follows. □

6 Order determination of functional autoregressive processes

While the last chapters all were concerned with random coefficient autoregressive processes, we will consider another process now. As mentioned in the introduction already, a generalization of the classical time series analysis is the functional time series analysis. Like before, we will also consider autoregressive processes. However, we do not restrict ourselves to processes of order one, but we will consider functional autoregressive processes of general order p and show how a reliable estimate of this order can be obtained.
We start with introducing the functional autoregressive model and show how to transform this model into an equivalent multivariate autoregressive model with non-standard noise to introduce our estimation procedure.
In the following, we will assume that we are given fully observed curves or curves obtained by smoothing curves with measurment errors or interpolating descrete observations. See Hall et al. (2006) who considered the case of sparse observations for how the methods can be adjusted for this case.

6.1 The functional autoregressive model

Let an arbitrary stationary functional time series be given that is rescaled to the interval $[0, 1]$, so that the process (1.1) turns into

$$Y_t(\tau), \quad t \in \mathbb{Z},\ \tau \in [0, 1]. \tag{6.1}$$

For convenience, we will drop the dependence on τ in the notation and just refer to the time series by Y_t, and will also do this for all other functions on the interval $[0, 1]$, where this does not lead to any confusion. Suppose that we are given observations $Y_1, \ldots, Y_n$ that are elements of the Hilbert Space $H = L^2([0, 1])$ equipped with the inner product $\langle x|y\rangle = \int_0^1 x(s)y(s)\mathrm{ds}$. This means that each $Y_t,\ t = 1, \ldots, n$, is a square integrable function satisfying $||Y_t||^2 = \langle Y_t|Y_t\rangle = \int_0^1 Y_t^2(s)\mathrm{ds} < \infty$. Introducing the following definition we can draw some more conclusions.

Definition 6.1.1. *For a functional process $Y_t(\tau),\ \tau \in [0, 1]$, we define for $r \in \mathbb{N}$ the number*

$$\nu_r(Y_t) = \boldsymbol{E}\left[||Y_t||^r\right]^{\frac{1}{r}}.$$

Every process $Y_t(\tau)$ with $\nu_1(Y_t) < \infty$ possesses a mean curve $\mu(\tau) = \mathbf{E}\left[Y_t(\tau)\right]$ and every process $Y_t(\tau)$ with $\nu_2(Y_t) < \infty$ possesses a covariance operator $C(x) = \mathbf{E}\left[\langle Y_t - \mu|x\rangle (Y_t - \mu)\right]$.

This operator is a kernel operator and can also be written as $C(x)(\tau) = \int_0^1 c(\tau, s)x(s)\mathrm{d}s$ with $c(\tau, s) = \mathbf{Cov}\left[Y_t(\tau), Y_t(s)\right]$ and it admits the spectral decomposition

$$C(x) = \sum_{\ell=1}^{\infty} \lambda_\ell \langle v_\ell | x \rangle \, v_\ell$$

where $\lambda_\ell,\ \ell \in \mathbb{N}$, are the eigenvalues in strictly decreasing order and $v_\ell,\ \ell \in \mathbb{N}$, are the corresponding normalized eigenfunctions, this means $C(v_\ell) = \lambda_\ell v_\ell$ and $||v_\ell|| = 1$. The eigenfunctions $v_\ell,\ l \in \mathbb{N}$, form an orthonormal basis of $L^2([0,1])$. Thus, Y_t allows for the Karhunen-Loéve representation

$$Y_t = \sum_{\ell=1}^{\infty} \langle Y_t | v_\ell \rangle \, v_\ell, \quad t \in \mathbb{Z} \tag{6.2}$$

where the coefficients $\langle Y_t | v_\ell \rangle$ are the functional principal components of Y_t. This representation will be of importance for the order selection that we propose later on.
In practice, having observations $Y_1, \ldots, Y_n$ at hand, both μ and C as well as its spectral decomposition are unknown and have to be estimated from the data. Let $\mu \equiv \hat{\mu}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} Y_t(\tau),\ \ \tau \in [0,1]$, be the estimated mean function. By Hörmann & Kokoszka (2012) (Lemma 4.1), $\hat{\mu}$ is $\sqrt{n}$-consistent for μ. This means that estimating the mean curve can be done in a separate step and we can assume that $\mathbf{E}\left[Y_t\right] = 0$. The covariance operator C can be estimated by the sample covariance operator $\hat{C}(x) \equiv \hat{C}_n(x) = \frac{1}{n}\sum_{t=1}^{n} \langle Y_t - \hat{\mu} | x \rangle \left(Y_t - \hat{\mu}\right)$. Again by Hörmann & Kokoszka (2012) (Theorem 2.1), this estimator is $\sqrt{n}$-consistent for C. For arbitrary, but typically small d, $\hat{C}$ can be used to compute estimated eigenvalues $\hat{\lambda}_{1,n} \equiv \hat{\lambda}_1, \ldots, \hat{\lambda}_{d,n} \equiv \hat{\lambda}_d$ and estimated eigenfunctions $\hat{v}'_{1,n}, \ldots, \hat{v}_{d,n}.'$ Since they are only determined up to the sign, we set $\hat{v}_1 \equiv \hat{v}_{1,n} = \mathrm{sign}\left(\langle \hat{v}'_{1,n} | \hat{v}_{1,n} \rangle\right) \hat{v}'_{1,n}, \ldots, \hat{v}_d \equiv \hat{v}_{d,n} = \mathrm{sign}\left(\langle \hat{v}'_{1,n} | \hat{v}_{d,n} \rangle\right) \hat{v}'_{d,n}$. Once more by Hörmann & Kokoszka (2012) (Theorem 3.2) these inherit $\sqrt{n}$-consistency for the true eigenvalues and eigenfunctions from $\hat{C}$. Following Hörmann & Kokoszka (2012), we let $\mathcal{L} = \mathcal{L}(H, H)$ be the set of bounded linear operators from H to H and define for $A \in \mathcal{L}$ the operator norm $||A||_{\mathcal{L}} = \sup_{||x \leq 1||} ||Ax||$. We will put some more assumptions on the process (6.1) and consider the functional AR(p)-model:

Definition 6.1.2. *If the process (6.1) has a mean function that is equal to zero everywhere and it is the unique stationary sequence (Hörmann & Kokoszka (2012), Example 2.1) satisfying the recursive equation*

$$Y_t(\tau) = \sum_{j=1}^{p} \phi_j\left(Y_{t-j}(\tau)\right) + \delta_t(\tau), \quad t \in \mathbb{Z},\ \tau \in [0,1] \tag{6.3}$$

it is called functional AR process of order p (FAR(p) process). Thereby, $\phi_j,\ j = 1, \ldots, p$, are linear Hilbert Schmidt integral operators such that the operator

$$\tilde{\phi} = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \ldots & I & 0 \end{pmatrix}$$

satisfies $||\tilde{\phi}||_{\mathcal{L}} < 1$ *and* $(\delta_t(\tau))_{t\in\mathbb{Z},}$ $\tau \in [0,1]$ *are i.i.d. error functions in* $L^2([0,1])$*. The operators* ϕ_j *can be represented as*

$$\phi_j(x)(\tau) = \int_0^1 \varphi_j(\tau, s)x(s)\,ds, \quad \textit{where} \int_0^1\int_0^1 \varphi_j^2(t,s)\,dtds < \infty.$$

Further, we assume $\nu_4(\delta_0) < \infty$*.*

This is not the most general definition of a FAR(p) process, however a version adapted to our situation. For a more general definition we refer, for example, to Bosq (2000), Chapter 3. The assumption $\nu_4(\delta_0) < \infty$ also implies that $\nu_4(Y_1) < \infty$.

Remark 6.1.3. *In the case* $p = 1$ *it can easily be seen that the process admits the expansion* $Y_t = \sum_{i=0}^{\infty} \phi^i \delta_{t-i}$*, where* ϕ^i *is the i-th iterate of the operator* ϕ*. A similar results holds true for* $p \geqslant 2$*, however, for notational convenience we abstain from stating it here.*

In the following, we assume that we have observations $Y_1, \ldots, Y_n$ at hand and want to determine the order p of the FAR(p) process $(Y_t)_{t\in\mathbb{Z}}$.
Equation (6.2) shows how a representation of the process $(Y_t)_{t\in\mathbb{Z}}$ in the orthonormal basis of the eigenfunctions of its covariance operator is given. Suppose we are given estimated eigenfunctions $\hat{v}_1, \ldots, \hat{v}_d$ with the correct sign as described before. We then can set up $Y_t^e = \sum_{\ell=1}^{d} \langle Y_t | \hat{v}_\ell \rangle \hat{v}_\ell$ $(t = 1, \ldots, n)$. For n growing to infinity the empirical principal component scores $\langle Y_t | \hat{v}_\ell \rangle$ converge to the true principal component scores and if additionally d is growing to infinity as well, the functions Y_t^e will be close to Y_t so that for finite d and n we can consider Y_t^e as a reasonable approximation to Y_t. In practice, one would use a principal component analysis to determine the representation of Y_t^e and would choose d such that a certain amount of variability of the data, for example 80%, is explained.
Introducing also empirical principal component scores for the operators ϕ_j $(j = 1, \ldots, p)$ and the noise functions δ_t $(t = 1, \ldots, n)$ the FAR(p) process transforms into a multivariate AR(p) process of dimension d (VAR(p) process). Usually, n observations of the functional model will give us n observations of the multivariate model as well. However, for our considerations, we would like to introduce $m = m(n)$. Given n observations of the functional model, we estimate its eigenfunctions and determine the multivariate model, but we transform only m functional observations into multivariate observations. This ensures that the error term coming into play because we use estimated eigenfunctions instead of true eigenfunctions decays fast enough. Therefore, the multivariate model is as follows:

Definition 6.1.4. *Let* $m = m(n) \overset{n\to\infty}{\longrightarrow} \infty$ *such that* $\frac{m^2}{n} \overset{n\to\infty}{\longrightarrow} 0$ *and d be arbitrary, for practicability such that the d first principal components of* Y *explain a sufficient high amount of variability. Further, let* $\hat{v}_1, \ldots, \hat{v}_d$ *be* $\sqrt{n}$*-consistent estimators of the eigenfunctions* $v_1, \ldots, v_d$ *of the covariance operator of* Y_t*. Suppose we are given observations* $Y_1, \ldots, Y_n$ *of the functional AR(p)-process (6.3), the representation of* Y_t *as a process similar to a multivariate AR(p)-process is*

$$\begin{pmatrix} \boldsymbol{Y}_{t,1} \\ \vdots \\ \boldsymbol{Y}_{t,d} \end{pmatrix} = \sum_{j=1}^{p} \Phi_j^e \begin{pmatrix} \boldsymbol{Y}_{t-j,1} \\ \vdots \\ \boldsymbol{Y}_{t-j,d} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_{t,1}^e \\ \vdots \\ \boldsymbol{\varepsilon}_{t,d}^e \end{pmatrix}, \quad t = 1, \ldots m$$

$$\boldsymbol{Y}_t = \sum_{j=1}^{p} \Phi_j^e \boldsymbol{Y}_{t-j} + \boldsymbol{\varepsilon}_t^e, \quad t = 1, \ldots m \tag{6.4}$$

where the vectors $\boldsymbol{Y}_t,\ t=1,\dots,n$ *consist of the empirical principal component scores of the functions* Y_t, *that means* $\boldsymbol{Y}_{t,i}=\langle Y_t|\hat{v}_i\rangle$, *while the vectors* $\boldsymbol{\varepsilon}_t$ *are the empirical principal component scores of the noise fuctions* δ_t, *that means* $\boldsymbol{\varepsilon}^e_{t,i}=\langle \delta_t|\hat{v}_i\rangle$, *and the matrices* Φ^e_j *consist of the empirical principal component scores of the operators* ϕ_j, *that means* $\Phi^e_{j,(i,k)}=\langle \phi_j\hat{v}_i|\hat{v}_k\rangle,\ j=1,\dots,p.$

We note that this process is not a standard VAR(p) process since the matrices Φ^e_j are random and dependent on each other and the noise terms $\boldsymbol{\varepsilon}^e_t,\ t=1,\dots,m$ are neither identically distributed, nor indenpendent from each other or independent from the observations $\boldsymbol{Y}_t,\ t=1,\dots,m$.
For the following considerations, we replace Φ^e_j by Φ_j, where each entry consists of the principal component score regarding the true basis of eigenfunctions, thus the remainder $\Phi_{j,(i,k)}-\Phi^e_{j,(i,k)}=\langle \phi_j\hat{v}_i|(\hat{v}_k-v_k)\rangle-\langle \phi_j\left(\hat{v}_i-v_i\right)|v_k\rangle$ is of order $\mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)$ and will therefore be included in the noise term, so that we consider the following model:

Definition 6.1.5. *Following the assumptions and notation of Definition 6.1.4, the modified representation of* Y *as a process similar to a multivariate AR(p)-process is given by*

$$\boldsymbol{Y}_t=\sum_{j=1}^{p}\Phi_j\boldsymbol{Y}_{t-j}+\boldsymbol{\varepsilon}_t,\quad t=1,\dots m, \tag{6.5}$$

where $\Phi_{j,(i,k)}=\langle \phi_j v_i|v_k\rangle$, $\boldsymbol{Y}_{t,i}=\langle Y_t|\hat{v}_i\rangle$, *and* $\boldsymbol{\varepsilon}_{t,i}=\langle \delta_t|\hat{v}_i\rangle+\sum_{k=1}^{d}\sum_{j=1}^{p}\langle \phi_j\hat{v}_i|(\hat{v}_k-v_k)\rangle-\langle \phi_j\left(\hat{v}_i-v_i\right)|v_k\rangle\boldsymbol{Y}_{t-j,k},\ i,k=1,\dots,d,\ t=1,\dots,m.$
With

$$\boldsymbol{X}_t=\begin{pmatrix}\boldsymbol{Y}_t\\ \boldsymbol{Y}_{t-1}\\ \vdots\\ \boldsymbol{Y}_{t-p+1}\end{pmatrix},\quad A=\begin{pmatrix}\Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p\\ I & 0 & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & I & 0\end{pmatrix},\quad \boldsymbol{\mathcal{E}}_t=\begin{pmatrix}\boldsymbol{\varepsilon}_t\\ 0\\ \vdots\\ 0\end{pmatrix}$$

the model (6.5) can be written as

$$\boldsymbol{X}_t=A\boldsymbol{X}_{t-1}+\boldsymbol{\mathcal{E}}_t,\quad t=1,\dots m$$

and with the noise terms $\boldsymbol{\mathcal{E}}_t=\boldsymbol{E}_t+\boldsymbol{U}_t+\frac{1}{\sqrt{n}}\boldsymbol{W}_t$ *splitted up as*

$$\boldsymbol{X}_t=A\boldsymbol{X}_{t-1}+\boldsymbol{E}_t+\boldsymbol{U}_t+\frac{1}{\sqrt{n}}\boldsymbol{W}_t,\quad t=1,\dots m, \tag{6.6}$$

where the vectors $\boldsymbol{E}_t=\left(e_{t,1},\dots,e_{t,d},0,\dots,0\right)^\top$, $\boldsymbol{U}_t=\left(u_{t,1},\dots,u_{t,d},0,\dots,0\right)^\top$, *and* $\boldsymbol{W}_t=\left(w_{t,1},\dots,w_{t,d},0,\dots,0\right)^\top$ *for* $t=1,\dots,m$ *and* $i=1,\dots,d$ *are given by:*

$e_{t,i}=\langle \delta_t|v_i\rangle,\quad$ *i.i.d. random variables,*

$u_{t,i}=\sum_{\ell=d+1}^{\infty}\sum_{j=1}^{p}\langle Y_{t-1}|v_\ell\rangle\langle \phi_j v_\ell|v_i\rangle$, *stationary random variables dependent on* $Y_{t-p},\dots,Y_{t-1}$

$$w_{t,i} = \sqrt{n}\sum_{\ell=1}^{d}\sum_{j=1}^{p}\left(\langle Y_{t-1}|v_\ell\rangle\langle\phi v_\ell|\hat{v}_i\rangle - \langle Y_{t-1}|\hat{v}_\ell\rangle\langle\phi\hat{v}_\ell|\hat{v}_i\rangle\right) + \sqrt{n}\sum_{\ell=d+1}^{\infty}\sum_{j=1}^{p}\langle Y_{t-1}|v_\ell\rangle\langle\phi_j v_\ell|\hat{v}_i - v_i\rangle$$
$$+ \sqrt{n}\sum_{\ell=1}^{d}\sum_{j=1}^{p}\langle\phi_j\hat{v}_i|(\hat{v}_\ell - v_\ell)\rangle - \langle\phi_j\left(\hat{v}_i - v_i\right)|v_\ell\rangle\boldsymbol{Y}_{t-j,\ell}, \quad \textit{bounded random variables}$$
$$\textit{dependent on } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m.$$

We observe that $\mathbf{E}\left[\boldsymbol{E}_t\right] = 0$ and $\mathbf{E}\left[\boldsymbol{U}_t\right] = 0$. By assumption, $v_l - \hat{v}_l = \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)$, and thus each entry $w_{t,i}$ is of order $\mathcal{O}_P(1)$. Defining

$$\boldsymbol{X}_t^{(1)} = A\boldsymbol{X}_{t-1}^{(1)} + \boldsymbol{E}_t, \qquad \boldsymbol{X}_t^{(2)} = A\boldsymbol{X}_{t-1}^{(2)} + \boldsymbol{U}_t, \qquad \boldsymbol{X}_t^{(3)} = A\boldsymbol{X}_{t-1}^{(3)} + \frac{1}{\sqrt{n}}\boldsymbol{W}_t,$$

we see that $\boldsymbol{X}_t^{(1)}$ is a standard VAR(1)-process since the vectors $\boldsymbol{E}_t\,(t = 1, \ldots, m)$ form a white noise and we can build up the process $\boldsymbol{X}_t$ as

$$\boldsymbol{X}_t = \boldsymbol{X}_t^{(1)} + \boldsymbol{X}_t^{(2)} + \boldsymbol{X}_t^{(3)}, \tag{6.7}$$

hence, it is a VAR(1)-process overlied with some noise terms. For the term $\boldsymbol{X}_t^{(1)}$ standard methods can be applied, whereas the term $\boldsymbol{X}_t^{(2)}$ adds significant statistical complexity since it is a VAR process whose error terms are dependent and form a process that is similar to a VAR process itself. It does not tend to zero, however, its size can be controlled by choosing the dimension d of the multivariate process. The higher the dimension, the smaller the "error" $\boldsymbol{X}_t^{(2)}$. Nevertheless, it should be pointed out that we do not need the dimension d to grow to infinity to obtain statistical reliable results since for fixed d this term is the same for all possible choices of the AR order and thus does not influence our procedure negatively. We note that by $\boldsymbol{X}_t^{(3)}$ a complex dependence structure comes into play. This is due to the fact that we use empirical principal component scores for the representation of the functional Y_t as the multivariate $\boldsymbol{Y}_t$. However, this dependence structure turns out to be asymptotically negligible since the terms will tend to zero with order $\sqrt{n}$. Because all following considerations with the multivariate process $\boldsymbol{Y}_t$ or $\boldsymbol{X}_t$, respectively, are made with m observations, there is not added any asymptotic bias term to our procedure.
It should be noted that the coefficient matrices Φ_j^e of the VAR process $\boldsymbol{Y}_t$ change everytime we add one observation. However, because we could asymtotically replace the matrices by the matrices Φ_j that depend only on the "true" eigenfunctions of the functional process we could circumvent this fact. However, the values of $\boldsymbol{W}_t\,(t = 1, \ldots, m)$ are random and adding one observation of the functional process Y_t will alter all of these random variables. In contrast, the random variables $\boldsymbol{E}_t$ and $\boldsymbol{U}_t\,(t = 1, \ldots, m)$ do not change if n changes. Nevertheless, varying d alters all random variables $\boldsymbol{U}_t$ and $\boldsymbol{W}_t\,(t = 1, \ldots, m)$.

6.2 Determination of the order

For the remainder of the chapter, we use the notation of Definition 6.1.5 and the discussion thereafter. Let the largest order of the process $P \in \mathbb{N}$ to test be arbitrary. We successively fit a VAR(q)-process to the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ for $q = 0, \ldots, P$ and determine the value of a loss function for each q. The minimal value will give us an estimate $\hat{p}$ of the order of the VAR(p)-process $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ and thus also of the order of the functional process $(Y_t)_{t\in\mathbb{Z}}$. We first have to introduce some more notation:

Definition 6.2.1. *We define the matrices*

$$\begin{aligned} B = B^\top &= \frac{1}{m-q}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)^\top \\ F &= \frac{1}{m-q}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top \\ G &= \frac{1}{m-q}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_{t+1} - \bar{\boldsymbol{X}}\right)^\top \\ \Sigma_q &= \boldsymbol{E}\left[\left(E_1 + U_1\right)\left(E_1 + U_1\right)^\top\right] \end{aligned}$$

and the standard estimators

$$\hat{A} = G^\top B^{-1\top} \tag{6.8}$$

$$\hat{\Sigma}_q = \frac{1}{m-q}\sum_{t=q+1}^{m}\left(\hat{\boldsymbol{\mathcal{E}}}_t - \bar{\boldsymbol{\mathcal{E}}}\right)\left(\hat{\boldsymbol{\mathcal{E}}}_t - \bar{\boldsymbol{\mathcal{E}}}\right)^\top \quad \textit{where } \hat{\boldsymbol{\mathcal{E}}}_t = \boldsymbol{X}_t - \hat{A}\boldsymbol{X}_{t-1},\ t = q,\dots,m. \tag{6.9}$$

The following results will only concern the choice of the order p of the FAR process for arbitrary dimension d of the auxiliary multivariate process. To estimate the order p correctly, no assumptions about the dimension d are necessary, we especially want to point out that we do not need the dimension d to grow to infinity. Asymptotically, every dimension d of the auxiliary multivariate process would give us the correct order p of the FAR process. However, in finite sample sizes, the choice of a meaningful dimension d is important and thus we will present some practical remarks concerning the choice of the dimension d subsequently.

With this, we can introduce the loss function that we would like to consider further:

Definition 6.2.2. *Let $\hat{\Sigma}_q$ be an estimate for the variance of the residuals of a VAR(q)-fit to the data $\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m$ as given in Definition 6.2.1. Let further $M = M(m) \overset{m\to\infty}{\longrightarrow} \infty$ such that $\frac{M}{m} \overset{m\to\infty}{\longrightarrow} 0$. Then, we define an objective function L by*

$$L(q) = d(q+1)\frac{M}{m} + \log_2\left(\det\hat{\Sigma}_q\right). \tag{6.10}$$

Then we obtain:

Lemma 6.2.3. *If a VAR-process of order $q \geqslant p$ is fitted to the data $\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m$ obtained from the transformation of the data $Y_1,\dots,Y_n$ stemming from a FAR-process of order p it holds true for the estimators of Definition 6.2.1 that*

$$\begin{aligned} \hat{A} &= A + \mathcal{O}_P\left(\frac{1}{\sqrt{m}} + \frac{\sqrt{m}}{\sqrt{n}}\right) \\ \text{and}\quad \hat{\Sigma}_q &= \frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top - \frac{1}{m-q}\left(\hat{A} - A\right)B\left(\hat{A} - A\right)^\top \qquad (6.11) \\ &= \Sigma_q + \mathcal{O}_P\left(\frac{1}{m}\right). \end{aligned}$$

The proof is delayed to Section 6.4. □

Lemma 6.2.4. *If a VAR-process of order $q < p$ is fitted to the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ obtained from the transformation of the data $Y_1, \ldots, Y_n$ stemming from a FAR-process of order p, it exists $\eta \in [0,1]$ (typically $\eta = 1$) such that*

$$\hat{\Sigma}_q = \hat{\Sigma}_p + \mathcal{O}_P\left(m^{\eta-1}\right) + \mathcal{O}_P\left(\frac{1}{\sqrt{m}}\right) \tag{6.12}$$

and $0 < \mathcal{O}_P\left(m^{\eta-1}\right)$.

For the proof we refer to Section 6.4. □

Now, we can introduce the estimator for the order of the process and state the main result.

Theorem 6.2.5. *Let the assumptions of this section and of Definition 6.2.2 hold. For $0 \leqslant q \leqslant P$ we successively fit a VAR-process of order q to the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ obtained from the transformation of the data $Y_1, \ldots, Y_m$ stemming from a FAR-process of order p. Then, the estimator*

$$\hat{p} = \underset{0 \leqslant q \leqslant P}{\operatorname{argmin}} L(q)$$

is consistent for p if $p \leqslant P$ and if not all observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ are identical.

Proof. In the way of proving the theorem we will follow Paulsen (1984) who considered standard VAR processes with unit roots. We first consider the case that we fit a VAR-process of higher than the true order to the data or of the true order, that means $p \leqslant q$. We define the parameter matrices $\Phi_{p+1}, \ldots, \Phi_q$ to be zero to obtain an VAR(q)-process from the VAR(q)-process $\boldsymbol{Y}_t$ ($t = 1, \ldots, m$). Following Definition 6.1.5 but for order q instead of p we transform this process into the VAR(1) process X_t ($t = 1, \ldots, m$). We note that for these processes $\Sigma_q = \Sigma_p$, so that Lemma 6.2.3 yields that $\hat{\Sigma}_q = \Sigma_p + \mathcal{O}_P\left(\frac{1}{m}\right)$ and $\hat{\Sigma}_p = \Sigma_p + \mathcal{O}_P\left(\frac{1}{m}\right)$ and hence

$$\hat{\Sigma}_q = \hat{\Sigma}_p + \mathcal{O}_P\left(\frac{1}{m}\right) = \hat{\Sigma}_p\left(I + \mathcal{O}_P\left(\frac{1}{m}\right)\right).$$

Therefore,

$$\begin{aligned}
L(q) &= d(q+1)\frac{M}{m} + \log_2\left(\det \hat{\Sigma}_p \det\left(I + \mathcal{O}_P\left(\frac{1}{m}\right)\right)\right) \\
&= L(p) + q\log_2\left(1 + \mathcal{O}_P\left(\frac{1}{m}\right)\right) + d(q-p)\frac{M}{m} \\
&= L(p) + \mathcal{O}_P\left(\frac{1}{m}\right) + d\underbrace{(q-p)}_{\geqslant 0}\frac{M}{m} \\
&= L(p) + d\underbrace{(q-p)}_{\geqslant 0}\frac{M}{m} + o\left(\frac{M}{m}\right),
\end{aligned}$$

thus for $p < q$

$$P\left\{L(q) > L(p)\right\} \stackrel{n\to\infty}{\longrightarrow} 1.$$

Let us now consider the case that we fit a VAR process of lower order than the true order to the data, that means $q < p$: Lemma 6.2.4 yields that

$$\hat{\Sigma}_q = \hat{\Sigma}_p + \mathcal{O}_P\left(m^{\eta-1}\right) + \mathcal{O}_P\left(\frac{1}{\sqrt{m}}\right)$$

with $\eta \in [0,1]$ (typically $\eta = 1$) and $0 < \mathcal{O}_P\left(m^{\eta-1}\right)$. Therefore,

$$\begin{aligned}
L(q) &= d(q+1)\frac{M}{m} + \log_2\left(\det\hat{\Sigma}_p \det\left(I + \mathcal{O}_P\left(m^{\eta-1} + \frac{1}{\sqrt{m}}\right)\right)\right)\\
&= L(p) + q\log_2\left(1 + \mathcal{O}_P\left(m^{\eta-1} + \frac{1}{\sqrt{m}}\right)\right) + d(q-p)\frac{M}{m}\\
&= L(p) + \underbrace{\mathcal{O}_P\left(m^{\eta-1}\right)}_{>0} + \mathcal{O}_P\left(\frac{1}{\sqrt{m}}\right) + \mathcal{O}\left(\frac{M}{m}\right),
\end{aligned}$$

thus for $q < p$

$$P\left\{L(q) < L(p)\right\} \stackrel{n\to\infty}{\longrightarrow} 1.$$

This completes the proof of the theorem. □

As previously indicated, the procedure as described in Theorem 6.2.5 provides an estimator for the order of the FAR(p) process for given dimension d of the auxiliary multivariate process, but it does not give any answer on how to choose d. One could think that the loss function (6.10) is dependent on p and d and hence p and d can be chosen simulataneously by comparing its values for various d and p. However, proceeding this way always leeds to choosing the largest dimension that is possible because the "true" d of the representation of the process $(Y_t)_{t\in\mathbb{Z}}$ in the basis of the eigenfunctions is infinity. Therefore, this method is not considered further.

The dimension d should be chosen such that a sufficient amount of the variability of the $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ is explained and such that the multivariate model is not too complex. Hence, it is reasonable to choose d as the smallest dimension that ensurs to explain, for example, 80% of the variability. Another possibility of choosing p and d simultaneously is the following: For given dimension d we determine the optimal order $\hat{p}_d$ and a measurement for the degree of seperation. This means, we ask how clear is the distinction between $\hat{p}_d$ and $\hat{p}_d \pm 1$. In a geometrical interpretation, we can determine the angle that the loss function exhibits in $\hat{p}_d$ as a function in p. The smaller this angle, the larger the degree of separation between $\hat{p}_d$ and its neighbors. For practicability, the dimension d can also be limited from below and above to exclude undesirable models that are obtained just because the parameter estimates for the multivariate model do not explain the paramters of the functional model well. This leads us to

Algorithm 6.2.6 (Highest degree of separation)**.** *Given observations* $Y_1, \dots, Y_n$ *of the functional process, transform them into multivariate observations* $\boldsymbol{Y}_1 \dots, \boldsymbol{Y}_m$ *of order* d *for* $d = D_{min}, \dots, D$ *where* D_{min} *is set to the smallest* d *so that at least* v_{min} *of the variability of the observations is explained and* D *is arbitrary. Determine the estimators* $\hat{p}_d$ *as well as the angle* α_d *the loss function exhibits in* $\hat{p}_d$. *If* $\hat{p}_d$ *is equal to the smallest or largest order tested, mirror the function in this point to determine the angle. Choose*

$$\hat{d} = \underset{D_{min} \leqslant d \leqslant D}{\operatorname{argmin}} \ \alpha(d) \qquad \textit{and accordingly} \qquad \hat{p} = \hat{p}_{\hat{d}}. \qquad \square$$

6.3 Finite sample behavior

To evaluate the finite sample behavior of our method, we perform a simulation study. Thereafter, we also apply the method to real data. But first, we have to specify the loss function (6.10) and a way how to choose the dimension d of the auxiliary multivariate process.

For the loss function we will focus on the Minimum Description Length principle (MDL). It is a statistical method for selecting the best fitting model from a class of candidate models that has been applied with great success to tackle different statistical problems, including structural break estimation for autoregressive processes. The MDL principle uses the code length for encoding the data as a means for comparing model complexity and data fidelity. The best fitting model is defined as the one that gives the shortest code length. There are different forms of MDL and we focus on the so-called two part MDL. The idea is to split the data $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_m$ into two parts: the fitted model $\hat{\theta}$ plus the corresponding residuals given the fitted model $\hat{\boldsymbol{\mathcal{E}}}|\hat{\theta}$. Denoting the code length of an object x by $C(x)$ we obtain $C(\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n) = C(\hat{\theta}) + C(\hat{\boldsymbol{\mathcal{E}}}|\hat{\theta})$. It is shown by Rissanen (1989) that the last term is equivalent to the negative of the log likelihood of the data, so MDL can be seen as a penalized likelihood method, where the penalty is derived from $C(\hat{\vartheta})$. By Lee (2001) it follows that in our case the objective function is

$$MDL(d, p, \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_m) = \frac{1}{4}\frac{\log m}{m}\left(4 + 3d + 2pd^2 + d^2\right) + \frac{\log p}{m} + \log\left(\det \hat{\Sigma}_{ML,q}\right). \tag{6.13}$$

In the presence of normally distributed errors, the ML estimator for A and the LS estimator (6.8) coincide (see Lütkepohl (2005), Section 3.4.3), and hence the loss function is exact when using Σ_q as given in Equation (6.9). In other cases, we can see it as an approximation. Even though we focus on the MDL, we also consider the AIC criterion (see Akaike (1974)) since it is very commonly used. However, the proof of consistency does not cover this criterion. For VAR processes it can even be shown that it is not consistent, so we expect that it is asymptotically not consistent here as well.

Regarding the choice of the dimension d of the auxiliary multivariate process, our presentation concentrates on the results of Algorithm 6.2.6 (highest degree of separation) since this has turned out to provide better results than the way with asking for a certain amount of variability to be explained. But we will also state some results concering this alternative.

6.3.1 A simulation study

The set-up for this simulation study consists of $N = 15$ or $N = 60$ cubic B-spline functions $z_1, \ldots, z_N$ on the unit interval $[0, 1]$, which together determine the space $H = \text{span}\{z_1, \ldots, z_N\}$. We note that each element x in H has the representation $x(\tau) = \sum_{\ell=1}^{M} c_\ell v_\ell(\tau)$ with coefficients $c_1, \ldots, c_M$. Innovations are defined in different ways: using $N = 15$, innovations are created by

$$\delta_t(\tau) = \sum_{\ell=1}^{15} D_{t,\ell} z_\ell(\tau),$$

where $(D_{t,1}, \ldots, D_{t,15})^\top$ are i.i.d. random vectors with independent either t (with four degrees of freedom) or standard normally distributed components. $N = 60$ is used to represent a standard brownian motion or a standard brownian bridge, respectively, in the space H. If $\phi : H \to H$ is a linear operator, then

$$\phi(x) = \sum_{\ell=1}^{N} c_\ell \phi(z_\ell) = \sum_{\ell_1=1}^{N} \sum_{\ell_2=1}^{N} c_{\ell_1} \langle \phi(z_{\ell_1}) | z_{\ell_2} \rangle z_{\ell_2} = (\boldsymbol{\phi c})^\top (z_1, \ldots, z_N)$$

where $\boldsymbol{\phi}$ is the matrix with entries $\boldsymbol{\phi}_{i,j} = \langle \phi z_i | z_j \rangle$. The linear operator ϕ can thus be represented by a $N \times N$ matrix that operates on the coefficients in the basis function representation of the curves.

For sample sizes $n = 100, 200, 1000$ we simulate realizations of the following functional processes over the interval $[0, 1]$ with the noise terms δ_t as given above.

$$Y_t(\tau) = \phi_1 Y_{t-1}(\tau) + \phi_2 Y_{t-2}(\tau) + \delta_t(\tau) \tag{6.14}$$

$$Y_t(\tau) = \phi_3 Y_{t-1}(\tau) + \phi_4 Y_{t-2}(\tau) + \phi_5 Y_{t-3}(\tau) + \phi_6 Y_{t-4}(\tau) + \delta_t(\tau) \tag{6.15}$$

$$Y_t(\tau) = \phi_1 Y_{t-2}(\tau) + \phi_2 Y_{t-4}(\tau) + \delta_t(\tau) \tag{6.16}$$

where the operators $\phi_1, \ldots, \phi_6$ are represented by

- ϕ_1: for model I $\boldsymbol{\phi}^{(A)}$ and for model II $\boldsymbol{\phi}^{(C)}$
- ϕ_2: for model I $\boldsymbol{\phi}^{(B)}$ and for model II $\boldsymbol{\phi}^{(C)}$
- ϕ_3: for model I $-0.4\boldsymbol{\phi}^{(A)}$ and for model II $-0.4\boldsymbol{\phi}^{(C)}$
- ϕ_4: for model I $0.2\boldsymbol{\phi}^{(B)}$ and for model II $0.2\boldsymbol{\phi}^{(C)}$
- ϕ_5: for model I $-0.4\boldsymbol{\phi}^{(A)}$ and for model II $-0.4\boldsymbol{\phi}^{(C)}$
- ϕ_6: for model I $0.8\boldsymbol{\phi}^{(B)}$ and for model II $0.8\boldsymbol{\phi}^{(C)}$

where we set $\boldsymbol{\phi}^{(A)} = \text{diag}\left(\frac{1}{4}, \ldots, \frac{1}{4}\right)$, $\boldsymbol{\phi}^{(B)}$ is chosen to have $\frac{1}{3}$ on the diagonal and $\frac{1}{4}$ on the first lower off-diagonal and zero else, and the matrix of a cosinues taper is set to be $\boldsymbol{\phi}^{(C)}_{i,j} = \cos(2\pi((i + j)/N - 1/2))/N$.

We set $m = \lfloor (\log(n + 50)^{3.25}) \cdot n^{0.49} \cdot 0.91 - 10 \rfloor$ to determine the number of multivariate observations $\boldsymbol{Y}_t$. All results cited in the following are based on 10 000 repetitions. For each process, a burn-in period of 80 is used. To compare our method to the existing one proposed by Kokoszka & Reimherr (2012), we also perform the simulations for this method (abbreviated KR).

Sample Size		n=100				n=200				n=1000			
Noise Type		T	N	BM	BB	T	N	BM	BB	T	N	BM	BB
KR Method	Model I	19	20	94	93	7	7	95	95	0	0	95	94
	Model II	85	86	93	91	93	92	88	82	95	94	20	4
Alg. 6.2.6 & MDL	Model I	97	98	92	99	100	100	100	100	100	100	100	100
	Model II	57	56	47	80	64	61	87	99	100	100	100	100
Var$\geqslant$ 80% & MDL	Model I	98	99	57	80	100	100	69	93	100	100	80	99
	Model II	64	61	77	76	90	90	86	81	79	80	77	78
Alg. 6.2.6 & AIC	Model I	89	91	90	97	100	100	99	100	100	100	100	100
	Model II	46	45	75	85	89	90	98	100	100	100	100	100
Var$\geqslant$ 80% & AIC	Model I	88	89	32	72	97	99	33	81	99	100	34	85
	Model II	82	76	63	56	95	53	98	59	77	46	39	44

Table 6.1: Percentage of correct detection of the order of the FAR(2) process (6.14)

Sample Size		n=100				n=200				n=1000			
Noise Type		T	N	BM	BB	T	N	BM	BB	T	N	BM	BB
KR Method	Model I	45	55	34	34	33	21	72	74	2	2	91	93
	Model II	8	9	40	26	39	40	85	71	81	81	81	75
Alg. 6.2.6 & MDL	Model I	78	78	70	82	100	100	98	100	100	100	100	100
	Model II	18	15	7	25	41	39	15	40	99	98	99	100
Var$\geqslant$ 80% & MDL	Model I	56	55	55	78	98	97	68	92	100	100	79	99
	Model II	2	1	64	69	1	1	81	89	91	89	88	96
Alg. 6.2.6 & AIC	Model I	56	55	72	66	99	100	96	99	100	100	100	100
	Model II	18	17	53	52	65	66	85	96	99	100	99	100
Var$\geqslant$ 80% & AIC	Model I	46	45	54	66	97	98	41	78	100	100	43	84
	Model II	6	5	52	51	37	37	85	59	99	100	46	57

Table 6.2: Percentage of correct detection of the order of the FAR(4) process (6.15)

Sample Size		n=100				n=200				n=1000			
Noise Type		T	N	BM	BB	T	N	BM	BB	T	N	BM	BB
KR Method	Model I	25	25	34	59	11	11	33	34	0	1	26	28
	Model II	6	6	5	5	5	5	5	5	7	7	7	4
Alg. 6.2.6 & MDL	Model I	92	94	89	97	100	100	100	100	100	100	100	100
	Model II	14	17	26	50	12	15	41	65	92	92	100	100
Var$\geqslant$ 80% & MDL	Model I	89	89	60	79	100	100	70	92	100	100	80	99
	Model II	35	34	65	63	63	64	77	76	77	78	76	76
Alg. 6.2.6 & AIC	Model I	82	83	86	93	99	100	99	100	100	100	100	100
	Model II	26	26	50	69	39	39	74	96	100	100	99	100
Var$\geqslant$ 80% & AIC	Model I	69	49	40	70	97	97	42	63	99	100	43	85
	Model II	29	16	57	57	47	46	62	65	56	55	51	53

Table 6.3: Percentage of correct detection of the order of the sparse FAR(4) process (6.16)

Tables 6.1 through 6.3 show the percentage of correct order detection for various models. We note that for the setting that Kokoszka & Reimherr (2012) considered, namely basically model I with brownian bridge errors, our simulation results and theirs coincide. In general, the method we propose works very well, in some settings it also outperforms the method introduced by Kokoszka & Reimherr (2012) that espacially cannot detect the process (6.16), what is clear because in the sequential testing the method detects that there is no dependence to the lag three in this process. However, our method can also detect these processes very well.

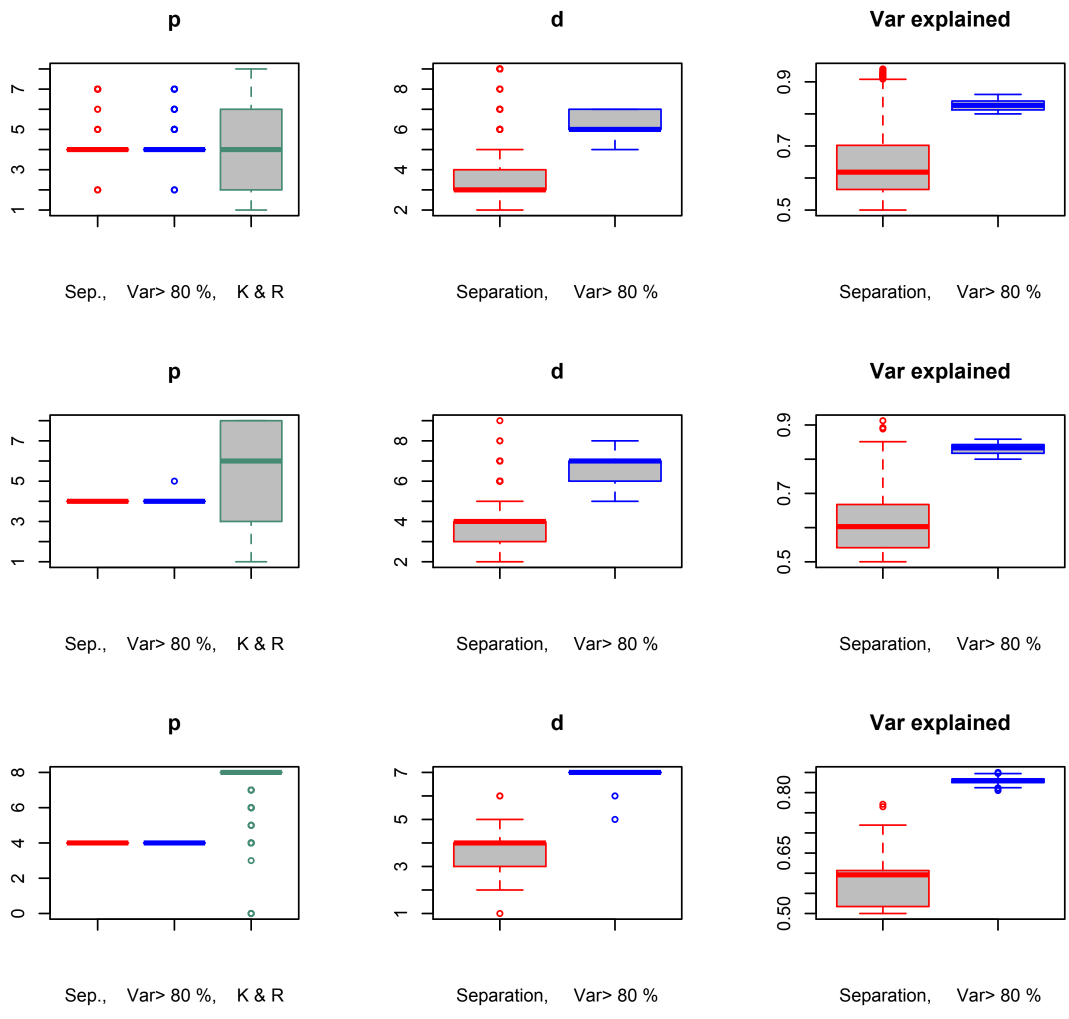

Figure 6.1: Orders chosen (p and d) and the explained variance for the FAR(4) process (6.16) with model I and t-distributed errors by several methods, from top to bottom $n = 100,\ 200,\ 1000$

As it can be seen in Tables 6.1 through 6.3, the Algorithm 6.2.6 works very well together with the MDL criterion and the model I whose parameter matrices basically have entries along the diagonals, for small sample sizes already. For model II that has more complicated structures in the parameter matrices, a slightly larger sample size is needed to produce results of the same quality. Also, with the AIC criterion, the quality is a bit lower. In

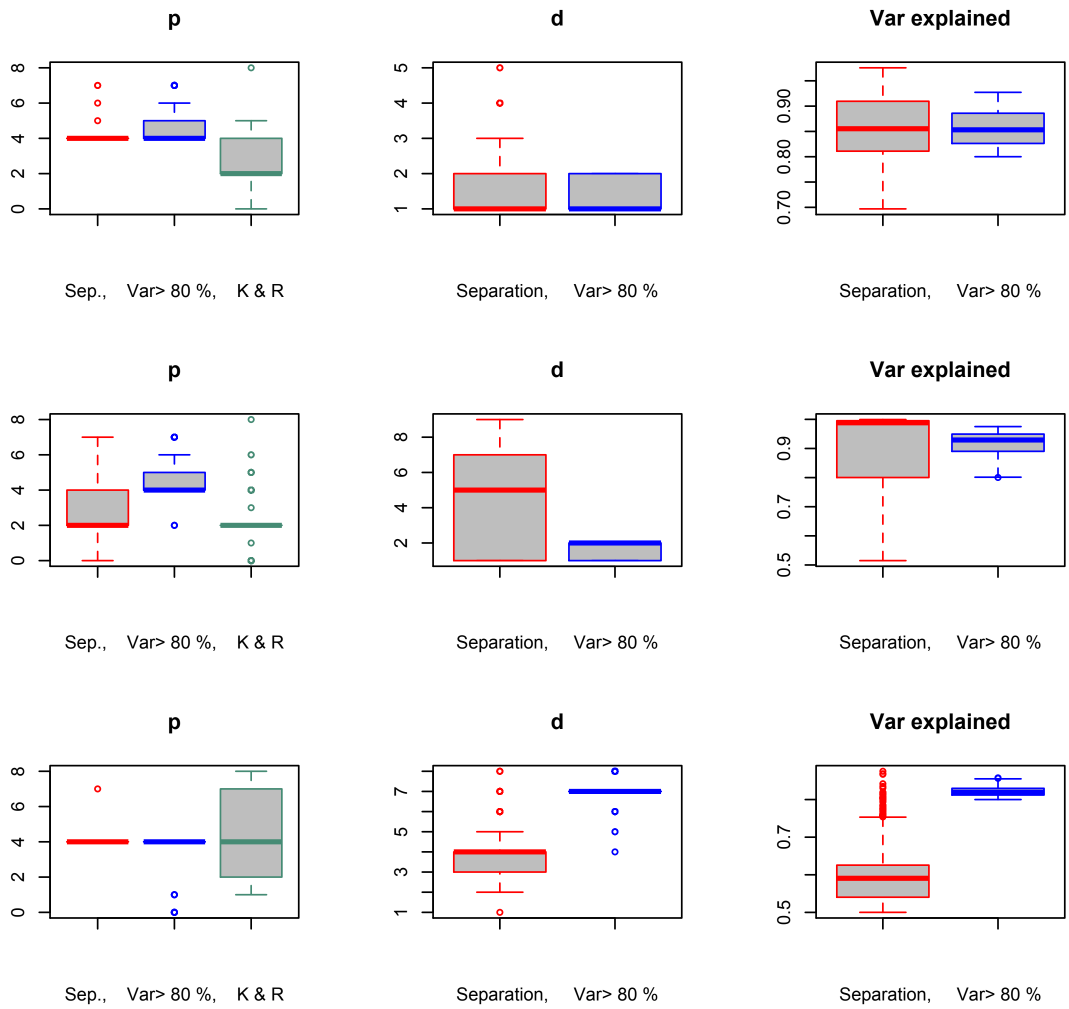

Figure 6.2: Orders chosen (p and d) and the explained variance for the FAR(4) process (6.16) with $n = 200$ by several methods, top brownian motion rrros model I, middle model II, and bottom the FAR(4) process (6.15) with t-distributed errors and model I

general, especially for model II, the results are better for models with errors that are independent sequences of dependent functions (brownian motion and brownian bridge) in contrast to independent sequences of independent random variables (t_4 and normal distribution). This is clear, since errors of the latter type destroy more of the structure of the time series. We want to point out that if we speak of (in)dependent errors in the following, we mean that there is (in)dependence between different points τ_1 and τ_2 of one curve. Of course, two different error curves are always independent. All of these observations basically hold for all three processes under consideration, even though the quality is better for models with less parameters.

If our method does not choose the right order, it tends to give a larger order, whereas the KR-algorithm over- and underestimates the order and in some cases it does not detect any order and chooses the largest order possible. This is displayed in Figures 6.1 and

6.2. The variance that is explained by the number of principal components d chosen in the "separation" case is smaller for i.i.d. errors then for dependent errors, whereas in the "variance> 80%" case it is smaller for dependent errors. This goes along with a smaller variance that is explained by the results of the "separation" than the "variance> 80%" for i.i.d. errors and about the same degree explanation for dependent errors. The dimension d chosen by Algorithm 6.2.6 is much smaller for the brownian motion and bridge errors than for the independent errors. This is because in the former case already one principal component explaines about 80% of the variability of the data in contrast to 30% in the latter case. Furthermore, it can be seen (Figure 6.3) that the dimension d that is chosen by the algorithm converges to four. For small sample sizes, it tends to choose a lower dimension, but it also sometimes chooses larger dimensions.

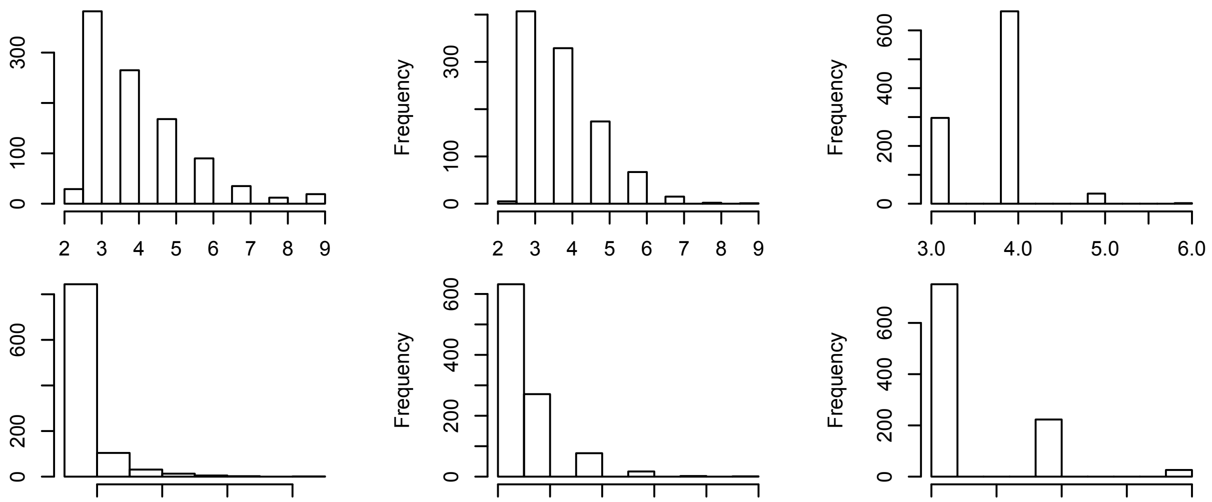

Figure 6.3: Distribution of dimensions d chosen with the Algorithm 6.2.6 for the FAR(2) process (6.14) for $n = 100,\ 200,\ 1000$ and t-distributed errors (plots in first row) and brownian motion errors (plots in second row)

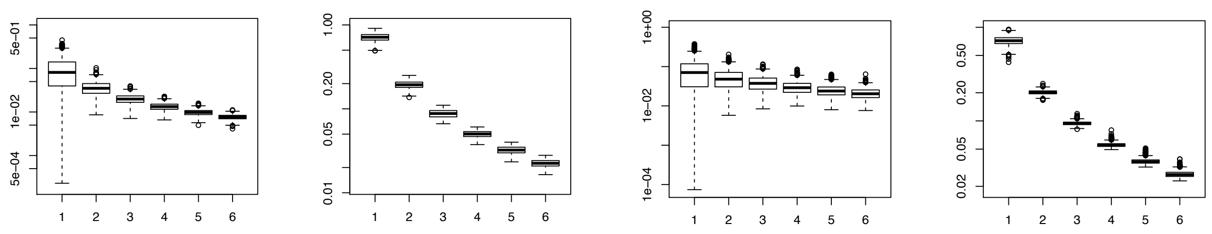

Figure 6.4: Relative error of the estimated parameter matrices depending on the dimension d for the FAR(2) process (6.14) (first and second matrix each), model I with normally distributed errors (left) and brownian motion errors (right)

The parameter matrices are also estimated well, as indicated by Figures 6.4 and 6.5. For small d it is clear that the errror is large, but for the d that is finally chosen by the algortihms the relative error of the parameter matrices is reasonable small. The second matrix tends to be harder to be estimated and in general the more complicated matrices of model II are harder to estimate than the more simple matrices of model I.

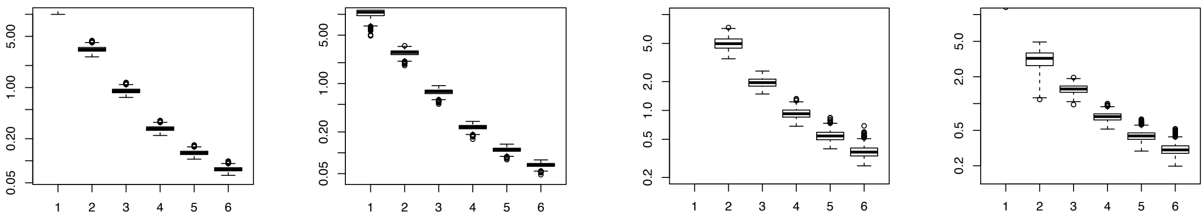

Figure 6.5: Relative error of the estimated parameter matrices depending on the dimension d for the FAR(2) process (6.14) (first and second matrix each), model II with normally distributed errors (left) and brownian motion errors (right)

6.3.2 Real data example

We consider two data sets: one set of annual australian mortality data from 1900 through 2003 and one set containing half hourly temperature and pollution data in Graz and the surrounding area (Klagenfurt) from 01.10.2010 through 31.03.2011. For our analysis we exclude the data from 31.12.2010 and 01.01.2011 due to some large outliers in the pollution data. Both of these data sets have been frequently used for statistical purpose, see, for example, Stadtlober et al. (2008) for an extensive discussion of the temperature and pollution data set. We first note the mortality data consists of logarithms of moratality rates and that following previous studies we use the root of the pollution data to account for few relatively large data point. First of all, we represent the data by a basis of ten B-splines over the interval $[0, 1]$ thereby obtaining fully observed curves, substract the overall mean curve and adjusted the curves for seasonal behavior.

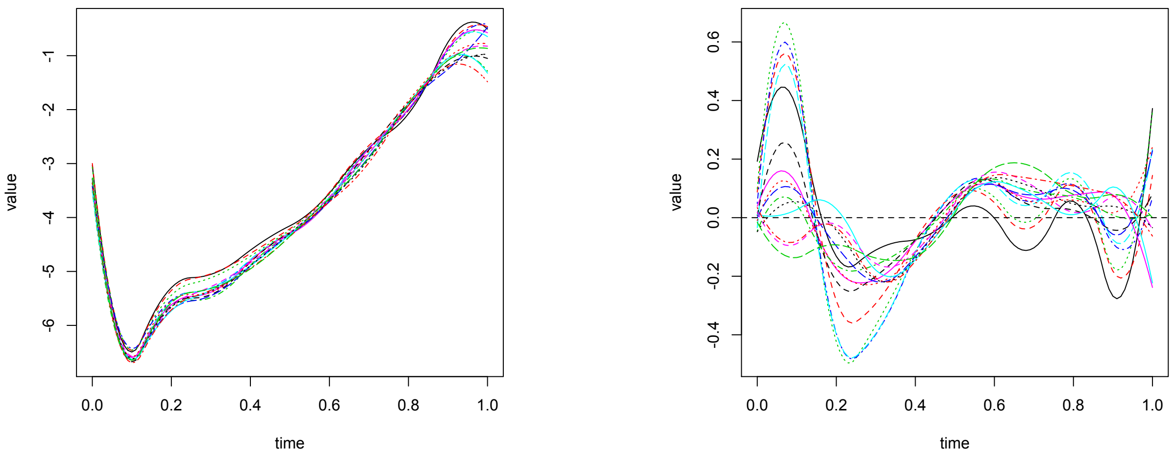

Figure 6.6: Mortality data set: first 15 curves, original (logarithmic) data (left) and mean adjusted data (right)

We start our considerations with a graphical analysis of the data. We have plotted the first 15 curves of each data set together with the mean adjusted curves, see Figures 6.6 and 6.7. The temperature curves follow the usual pattern of temperatures and the curves of the pollution data show the peaks of the morning and evening rush-our.

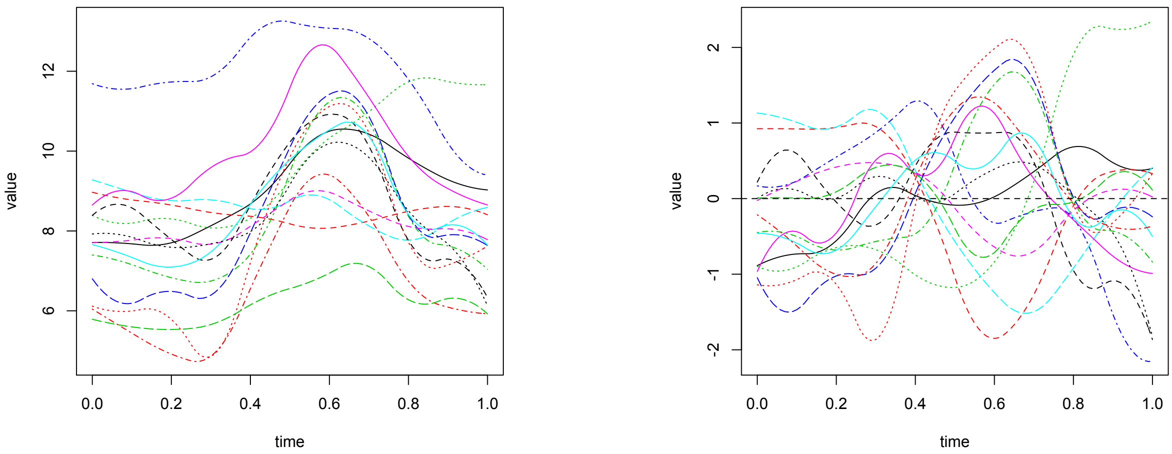

Figure 6.7: Temperature Klagenfurt data set: first 15 curves, original data (left) and mean adjusted data (right)

Figure 6.8 shows the correlation **Corr** $[Y_t(\tau_1), Y_{t+h}(\tau_2)]$ in the curves of the temperature data set between different points τ_1 and τ_2 in time of the same curve and with a shift of $h = 2$ or $h = 10$ curves in between. It can be seen that the data is not independent, but the correlation decays with the shift h between the curves increasing. Thus, it makes sense to fit an autoregressive model to the data.

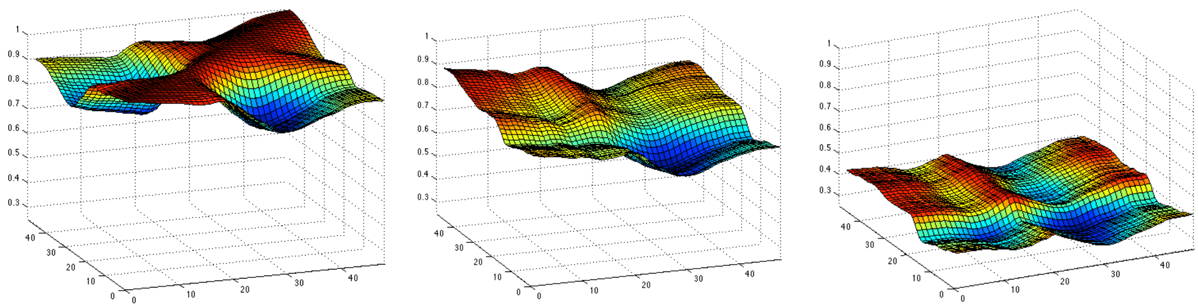

Figure 6.8: Temperature Klagenfurt data set: Autocorrelation in the curves (left) and with a time shift of $h = 2$ (middle) and $h = 10$ (right)

The models suggested by the various criteria (as MDL versus AIC and a fixed variability that has to be explained versus the largest separation and the method of Kokoszka & Reimherr (2012)) are shown in Table 6.4. The parameters were set as: $D = 6$, $P = 5$ and $v_{min} = 0.7$. To compare with, we also give the results obtained by the multistage testing method of Kokoszka & Reimherr (2012). For most data sets, all suggested orders for p coincide, only the AIC criterion with at least 80% of the variability to explain differs.

In the next step, we fit a FAR(p) model of the order suggested by Algorithm 6.2.6 and detmine the estimated residuals. Figure 6.9 shows the correlation in the curve of the residuals after fitting a model to the data that is not negligible. Hence, it makes sense to

	Alg. 6.2.6		Var⩾80%		Alg. 6.2.6		Var⩾80%		
	MDL				AIC				KR
	p	d	p	d	p	d	p	d	p
Mortality Data	2	3	2	1	2	3	2	1	2
Pollution Data Graz	1	4	1	4	1	3	3	4	1
Pollution Data Klagenfurt	1	4	1	4	1	4	1	4	1
Temperature Data Graz	1	2	1	2	1	2	1	2	1
Temperature Data Kl.	1	4	0	2	1	5	4	2	2

Table 6.4: Suggested orders for the data sets

choose a model whose errors are not independent and we choose the errors for the model to follow a brownian bridge because a comparision with the respective figure that shows the estimated residuals from a model with brownian bridge errors indicates that this error type might not be a bad choice for these data.

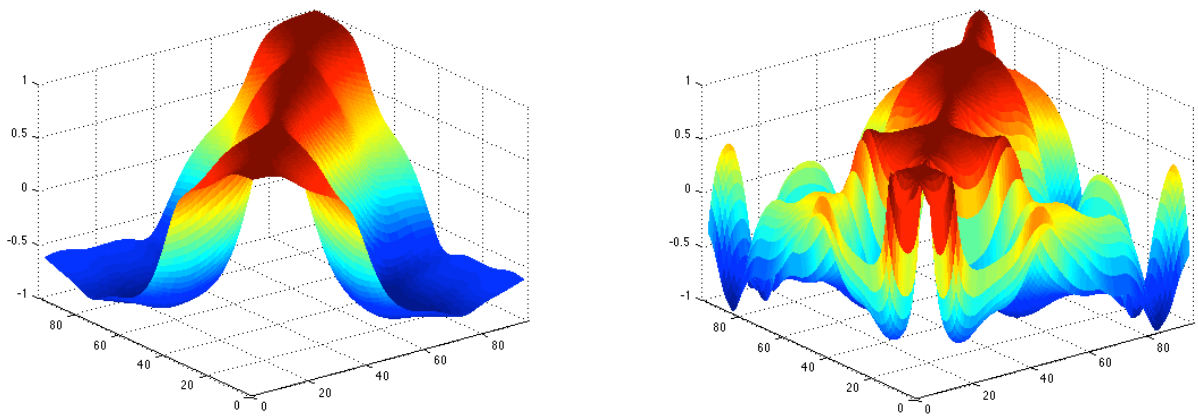

Figure 6.9: Autocorrelation in the error curves after fitting the suggested models to the data: Temperature data Klagenfurt (left) and mortality data (right)

Finally, we test the quality of the models by a prediction: we do not use the last five observations of the data sets for the determination of the order and estimation of the parameters, what leads to the same results as before, and predict five observations into the future. The prediction error is computed as the integral over the squared difference between the predicted and the true curves. The results (mean of 10 000 repetitions) are displayed in Table 6.5. For the mortality data set, the order two, that was chosen by all methods, leads to the smallest prediction error, for the temperature data set, the models of order zero and one, that were both chosen by the MDL criterion (but with the two different rules to choose d), lead to the smallest prediction error, hence, here as well as in the simulation study, the results obtained by using the MDL criterion, especially with Algorithm 6.2.6, are very reasonable and good so that the methods proposed here are a good method to determine the order of a FAR(p) process.

Mortality Data	$p=0,\, d=3$	$p=1,\, d=3$	$p=2,\, d=3$	$p=3,\, d=3$
relative Error	1.20	1.01	1.00	1.01
Temp. Klagenfurt Data	$p=0,\, d=2$	$p=1,\, d=4$	$p=1,\, d=5$	$p=2,\, d=3$
relative Error	1.00	1.01	1.48	1.16

Table 6.5: Errors for out-of-sample prediction for various models

6.4 Proofs

In this section, we state the proofs we have referred to before, but first we state three additional lemmas that will be used in these proofs.

Lemma 6.4.1. *Let $\boldsymbol{Z}_t = \boldsymbol{X}_t - \bar{\boldsymbol{X}}$. Then, using the assumptions and notation of Section 6.2 the following asymptotic results hold true, where the last one only holds true if not all observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ are identical.*

$$(I) \qquad N_m = \sum_{t=1}^{m-1} \boldsymbol{E}\left[\boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right] = \mathcal{O}\left(m\right)$$

$$(II) \qquad V_m' = N_m^{-\frac{1}{2}} \sum_{t=1}^{m-1} \boldsymbol{Z}_t (\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}})^\top = \mathcal{O}_P\left(1 + \frac{m}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)$$

$$(III) \qquad V_m = N_m^{\frac{1}{2}\top} \left(\sum_{t=1}^{m} \boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right)^{-1} N_m^{\frac{1}{2}} = \mathcal{O}_P\left(1\right)$$

Proof. For (I) consider

$$\begin{aligned}
\mathbf{E}\left[\boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right] &= \mathbf{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^\top - \boldsymbol{X}_t \bar{\boldsymbol{X}}^\top - \bar{\boldsymbol{X}} \boldsymbol{X}_t^\top + \bar{\boldsymbol{X}} \bar{\boldsymbol{X}}^\top\right] \\
&= \mathbf{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^\top\right] - \frac{1}{m} \sum_{s=1}^{m} \mathbf{E}\left[\boldsymbol{X}_t \boldsymbol{X}_s^\top\right] - \frac{1}{m} \sum_{s=1}^{m} \mathbf{E}\left[\boldsymbol{X}_s \boldsymbol{X}_t^\top\right] + \frac{1}{m^2} \sum_{s=1}^{m} \sum_{r=1}^{m} \mathbf{E}\left[\boldsymbol{X}_s \boldsymbol{X}_r^\top\right]
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t\right] &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A^i \mathbf{E}\left[\left(\boldsymbol{E}_{t-i} + \boldsymbol{U}_{t-i} + \frac{1}{\sqrt{n}} \boldsymbol{W}_{t-i}\right) \left(\boldsymbol{E}_{s-j} + \boldsymbol{U}_{s-j} + \frac{1}{\sqrt{n}} \boldsymbol{W}_{s-j}\right)^\top\right] A^{j\top} \\
&\leqslant \sum_{i=0}^{\infty} A^i C \sum_{j=0}^{\infty} A^{j\top} \\
&= \mathcal{O}(1)
\end{aligned}$$

since $\boldsymbol{E}_t$, $\boldsymbol{U}_t$, and $\boldsymbol{W}_t$ are assumed to have finite second moments. Hence, we obtain that $\mathbf{E}\left[\boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right] = \mathcal{O}(1)$ and therefore the assertion. For (II) we have

$$\begin{aligned}
\mathbf{E}\left[V_m'\right] = N_m^{-\frac{1}{2}} \sum_{t=1}^{m-1} \mathbf{E}\left[\boldsymbol{Z}_t (\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}})^\top\right] &= N_m^{-\frac{1}{2}} \sum_{t=1}^{m-1} \mathbf{E}\left[\boldsymbol{X}_t \boldsymbol{\mathcal{E}}_{t+1}^\top - \bar{\boldsymbol{X}} \boldsymbol{\mathcal{E}}_{t+1}^\top + \boldsymbol{X}_t \bar{\boldsymbol{\mathcal{E}}}^\top - \bar{\boldsymbol{X}} \bar{\boldsymbol{\mathcal{E}}}^\top\right] \\
&= N_m^{-\frac{1}{2}} \mathcal{O}(m) = \mathcal{O}\left(\frac{1}{\sqrt{m}}\right)
\end{aligned}$$

because

$$\mathbf{E}\left[\boldsymbol{X}_t\boldsymbol{\mathcal{E}}_{t+1}^\top\right] = \mathbf{E}\left[\boldsymbol{X}_t\left(\boldsymbol{E}_{t+1}^\top + \boldsymbol{U}_{t+1}^\top + \frac{1}{\sqrt{n}}\boldsymbol{W}_{t+1}^\top\right)\right] = 0 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}(1) = \mathcal{O}(1)$$

and

$$\begin{aligned}\mathbf{E}\left[\bar{\boldsymbol{X}}\boldsymbol{\mathcal{E}}_{t+1}\right] = \frac{1}{m}\sum_{s=1}^{m}\mathbf{E}\left[\boldsymbol{X}_s\boldsymbol{\mathcal{E}}_{t+1}^\top\right] &= \frac{1}{m}\sum_{s=1}^{m}\mathbf{E}\left[\boldsymbol{X}_s\left(\boldsymbol{E}_{t+1}^\top + \boldsymbol{U}_{t+1}^\top + \frac{1}{\sqrt{n}}\boldsymbol{W}_{t+1}^\top\right)\right]\\ &= \mathcal{O}\left(1 + \frac{1}{\sqrt{n}} + 1\right) = \mathcal{O}(1).\end{aligned}$$

To determine the variance we consider

$$\begin{aligned}&\mathbf{E}\left[\mathrm{vec}V_m'\left(\mathrm{vec}V_m'\right)^\top\right]\\ &= \left(I\otimes N_m^{-\frac{1}{2}}\right)\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\left(\boldsymbol{\mathcal{E}}_{t+1}\otimes\boldsymbol{Z}_t\right)\left(\boldsymbol{\mathcal{E}}_{s+1}\otimes\boldsymbol{Z}_s\right)^\top\right]\left(I\otimes N_m^{-\frac{1}{2}}\right)^\top\\ &= \left(I\otimes N_m^{-\frac{1}{2}}\right)\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\boldsymbol{\mathcal{E}}_{t+1}\boldsymbol{\mathcal{E}}_{s+1}^\top\otimes\boldsymbol{Z}_t\boldsymbol{Z}_s^\top\right]\left(I\otimes N_m^{-\frac{1}{2}}\right)^\top\\ &= \left(I\otimes N_m^{-\frac{1}{2}}\right)\left(\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\boldsymbol{\mathcal{E}}_{t+1}\boldsymbol{\mathcal{E}}_{s+1}^\top\otimes\boldsymbol{X}_t\boldsymbol{X}_s^\top\right] - \frac{2}{m}\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\sum_{r=1}^{m}\mathbf{E}\left[\boldsymbol{\mathcal{E}}_{t+1}\boldsymbol{\mathcal{E}}_{s+1}^\top\otimes\boldsymbol{X}_r\boldsymbol{X}_s^\top\right]\right.\\ &\qquad\left. + \frac{1}{m^2}\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\sum_{r=1}^{m}\sum_{l=1}^{m}\mathbf{E}\left[\boldsymbol{\mathcal{E}}_{t+1}\boldsymbol{\mathcal{E}}_{s+1}^\top\otimes\boldsymbol{X}_r\boldsymbol{X}_l\right]\right)\left(I\otimes N_m^{-\frac{1}{2}}\right)^\top.\end{aligned} \tag{6.17}$$

Consider the first sum

$$\begin{aligned}&\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\boldsymbol{\mathcal{E}}_{t+1}\boldsymbol{\mathcal{E}}_{s+1}^\top\otimes\boldsymbol{X}_t\boldsymbol{X}_s^\top\right]\\ &= \sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\left(\boldsymbol{E}_{t+1}^\top + \boldsymbol{U}_{t+1}^\top + \frac{1}{\sqrt{n}}\boldsymbol{W}_{t+1}^\top\right)\left(\boldsymbol{E}_{t+1}^\top + \boldsymbol{U}_{t+1}^\top + \frac{1}{\sqrt{n}}\boldsymbol{W}_{t+1}^\top\right)^\top\right.\\ &\qquad\left.\otimes\left(\boldsymbol{X}_t^{(1)} + \boldsymbol{X}_t^{(2)} + \boldsymbol{X}_t^{(3)}\right)\left(\boldsymbol{X}_s^{(1)} + \boldsymbol{X}_s^{(2)} + \boldsymbol{X}_s^{(3)}\right)^\top\right]\end{aligned} \tag{6.18}$$

Evaluating this term gives us numerous terms.

$$\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\boldsymbol{E}_{t+1}\boldsymbol{E}_{s+1}^\top\otimes\boldsymbol{X}_t^{(1)}\boldsymbol{X}_s^{(1)\top}\right] = \mathcal{O}(m)$$

according to Paulsen (1984) since this is just the standard case. However, the following proof also includes this case. All terms including at least one $\boldsymbol{W}_i$ or $\boldsymbol{X}_j^{(3)}$ are of order $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$. We consider the term

$$\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\boldsymbol{U}_{t+1}\boldsymbol{U}_{s+1}^\top\otimes\boldsymbol{X}_t^{(3)}\boldsymbol{X}_s^{(3)\top}\right].$$

In absolute value, each entry of this matrix is not larger than the absolute value of the entries of

$$\sum_{t=1}^{m-1}\sum_{s=1}^{m-1}\mathbf{E}\left[\left|\boldsymbol{U}_{t+1}\boldsymbol{U}_{s+1}^{\top}\boldsymbol{X}_t^{(3)}\boldsymbol{X}_s^{(3)\top}\right|\right].$$

By the argumentation of the proof of Lemma 6.4.2 this term is of order $\mathcal{O}(m)$. We note that for the case (2b) therein, the expression is of order $\mathcal{O}(m)$ because it reduces to the case $t+1=s$ and $s+1=t$ and all other cases can be evaluated similarly as in the proof there.
All other terms resulting out of Equation (6.18) can be evaluated similarly to this term since this is the most complex term. They contain one to four $\boldsymbol{E}$ instead of the $\boldsymbol{U}$, but for these the same proofs hold true since they are independent and not only asymptotically independent and feature the same characteristics. Hence, we obtain for Equation (6.17) by additionally applying the first result of this lemma

$$\mathbf{E}\left[\mathrm{vec}V_m'\left(\mathrm{vec}V_m'\right)^{\top}\right]=O\left(1+\frac{m}{\sqrt{n}}\right)$$

and therefore the assertion.
Regarding (III) we already know that $0<\mathbf{E}\left[\boldsymbol{Z}_t\boldsymbol{Z}_t^{\top}\right]=\mathcal{O}(1)$ as long as not all $Y_t\,(t=1,\ldots,n)$ are identical. The variance can be computed to be

$$\begin{aligned}
&\mathbf{Var}\left[\frac{1}{m}\sum_{t=1}^{m}\boldsymbol{Z}_t\boldsymbol{Z}_t^{\top}\right]\\
&=\frac{1}{m^2}\sum_{t=1}^{m}\sum_{s=1}^{m}\left(\mathbf{E}\left[\boldsymbol{Z}_t\boldsymbol{Z}_t^{\top}\boldsymbol{Z}_s\boldsymbol{Z}_s^{\top}\right]-\mathbf{E}\left[\boldsymbol{Z}_t\boldsymbol{Z}_t^{\top}\right]\mathbf{E}\left[\boldsymbol{Z}_s\boldsymbol{Z}_s^{\top}\right]\right)\\
&=\frac{1}{m^2}\sum_{t=1}^{m}\sum_{s=1}^{m}\left(\mathbf{E}\left[\left(\boldsymbol{Z}_t^{(1)}+\boldsymbol{Z}_t^{(2)}\right)\left(\boldsymbol{Z}_t^{(1)}+\boldsymbol{Z}_t^{(2)}\right)^{\top}\left(\boldsymbol{Z}_s^{(1)}+\boldsymbol{Z}_s^{(2)}\right)\left(\boldsymbol{Z}_s^{(1)}+\boldsymbol{Z}_s^{(2)}\right)^{\top}\right]\right.\\
&\qquad\left.-\,\mathbf{E}\left[\left(\boldsymbol{Z}_t^{(1)}+\boldsymbol{Z}_t^{(2)}\right)\left(\boldsymbol{Z}_t^{(1)}+\boldsymbol{Z}_t^{(2)}\right)^{\top}\right]\mathbf{E}\left[\left(\boldsymbol{Z}_s^{(1)}+\boldsymbol{Z}_s^{(2)}\right)\left(\boldsymbol{Z}_s^{(1)}+\boldsymbol{Z}_s^{(2)}\right)^{\top}\right]\right)+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

where $\boldsymbol{Z}_t$ is split up like $\boldsymbol{X}_t$ in Equation (6.7). Evaluating the double sum, using the definition $\boldsymbol{Z}_t=\boldsymbol{X}_t-\bar{\boldsymbol{X}}$ and introducing the $MA(\infty)$ representation of $\boldsymbol{X}_t$ gives us again numerous terms. As in the proof of Lemma 6.4.1, it is sufficient to look at the term

$$\begin{aligned}
&\frac{1}{m^2}\sum_{t=0}^{m}\sum_{s=0}^{m}\boldsymbol{X}_t^{(2)}\boldsymbol{X}_t^{(2)\top}\boldsymbol{X}_s^{(2)}\boldsymbol{X}_s^{(2)\top}\\
&\quad=\frac{1}{m^2}\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty}A^{i_1}\left(\mathbf{E}\left[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^{\top}A^{i_2\top}A^{i_3}\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^{\top}\right]\right.\\
&\qquad\qquad\left.-\mathbf{E}\left[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^{\top}\right]A^{i_2\top}A^{i_3}\mathbf{E}\left[\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^{\top}\right]\right)A^{i_4\top}
\end{aligned}$$

since all other terms are of the same or even smaller order. But with Lemma 6.4.2 this term is of order $\mathcal{O}\left(\frac{1}{m}\right)$ and hence

$$\frac{1}{m^2}\mathbf{Var}\left[\sum_{t=1}^{m}\boldsymbol{Z}_t\boldsymbol{Z}_t^{\top}\right]=\mathcal{O}\left(\frac{1}{m}\right).$$

Thus, it exists a matrix J such that $\frac{1}{m}\sum_{t=1}^{m} \boldsymbol{Z}_t\boldsymbol{Z}_t^\top \xrightarrow{\text{i.P.}} J$. J is positive definite, hence J^{-1} exists and is positive definite as well, so $\left(\frac{1}{m}\sum_{t=1}^{m} \boldsymbol{Z}_t\boldsymbol{Z}_t^\top\right)^{-1} = \mathcal{O}_P(1)$. Together with the fact that $\frac{1}{m}N_m = \mathcal{O}(1)$ the assertion follows and the proof of the lemma is complete. □

Lemma 6.4.2. *Under the assumptions and notation of Section 6.2 it holds*

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1}\Big(\boldsymbol{E}\Big[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^\top {A^{i_2}}^\top A^{i_3}\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^\top\Big] \\ -\boldsymbol{E}\big[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^\top\big] {A^{i_2}}^\top A^{i_3}\boldsymbol{E}\big[\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^\top\big]\Big) {A^{i_4}}^\top = \mathcal{O}(m).$$

Proof. Without loss of generalization, we assume $p = 1$ for the proof. We first note that for $\rho < 1$

$$\sum_{t=0}^{m}\sum_{s=0}^{t}\rho^{|t-s|} = 2\sum_{t=0}^{m}\sum_{s=0}^{m}\rho^{t-s} + \mathcal{O}(m) = 2\sum_{t=0}^{m}\rho^t\frac{1-\rho^{-t-1}}{1-\rho^{-1}} + \mathcal{O}(m) = \mathcal{O}(m).$$

Let us have a closer look at

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1}v_{k_1}v_{k_2}^\top {A^{i_2}}^\top A^{i_3}v_{k_3}v_{k_4}^\top {A^{i_4}}^\top \\ \Big(\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right] \\ -\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\Big)$$

and separate five cases:

1. *all four indices are equal:* $i_1 = i_2 = t-s+i_3 = t-s+i_4$:
   like in the standard case we obtain that

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1}v_{k_1}v_{k_2}^\top {A^{i_2}}^\top A^{i_3}v_{k_3}v_{k_4}^\top {A^{i_4}}^\top = \mathcal{O}(m).$$

2. *always exactly two indices are equal:*

   a) $i_1 = i_2$, $i_3 = i_4$:

$$(\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right] \\ -\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]) \\ \leqslant ||\phi||_{\mathcal{L}}^{|t-s+i_1-i_3|}K\nu_2(\delta_0)$$

   by Lemma 6.4.3 and hence

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1}v_{k_1}v_{k_2}^\top {A^{i_2}}^\top A^{i_3}v_{k_3}v_{k_4}^\top {A^{i_4}}^\top ||\phi||_{\mathcal{L}}^{|t-s+i_1-i_3|}K\nu_2(\delta_0) = \mathcal{O}(m).$$

b) $t-i_1 = s-i_3$, $t-i_2 = s-i_4$ and $t-i_1 = s-i_4$, $t-i_2 = s-i_3$:
Easy computations yield that

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1} v_{k_1} v_{k_2}^{\top} A^{i_2\top} A^{i_3} v_{k_3} v_{k_4}^{\top} A^{i_4\top} = \mathcal{O}(m).$$

3. *the index relating to the last point in time and the index relating to the second but last point in time are different: $t_1 \leqslant t_2 \leqslant t_3 < t_4$:*

$$\mathbf{E}\left[\langle Y_{t_1}|v_{\ell_1}\rangle\langle Y_{t_2}|v_{\ell_2}\rangle\langle Y_{t_3}|v_{\ell_3}\rangle\langle Y_{t_4}|v_{\ell_4}\rangle\right] \leqslant ||\phi||_{\mathcal{L}}^{t_4-t_3} K\nu_4(\delta_0)^4$$

by Lemma 6.4.3 and for $a = 1,2,3$

$$\mathbf{E}\left[\langle Y_{t_a}|v_{\ell_1}\rangle\langle Y_{t_4}|v_{\ell_4}\rangle\right] = ||\phi||_{\mathcal{L}}^{t_4-t_a} K\nu_2(\delta_0)^2$$

and hence

$$\begin{aligned}&\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1} v_{k_1} v_{k_2}^{\top} A^{i_2\top} A^{i_3} v_{k_3} v_{k_4}^{\top} A^{i_4\top}\\ &\qquad\Big(\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\\ &\qquad\quad - \mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\Big)\\ &= \mathcal{O}(m).\end{aligned}$$

4. *the indices relating to the last two points in time are equal, the two remaining are different:*
Easy calculations yield that

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1} v_{k_1} v_{k_2}^{\top} A^{i_2\top} A^{i_3} v_{k_3} v_{k_4}^{\top} A^{i_4\top} = \mathcal{O}(m).$$

5. *the indices relating to the last two points in time and one other index are equal, the remaining is different*
Easy calculations yield that

$$\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1} v_{k_1} v_{k_2}^{\top} A^{i_2\top} A^{i_3} v_{k_3} v_{k_4}^{\top} A^{i_4\top} = \mathcal{O}(m).$$

By plugging these results together we obtain that

$$\begin{aligned}&\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1} v_{k_1} v_{k_2}^{\top} A^{i_2\top} A^{i_3} v_{k_3} v_{k_4}^{\top} A^{i_4\top}\\ &\qquad\Big(\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\\ &\qquad\quad - \mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\Big)\\ &= \mathcal{O}(m).\end{aligned}$$

For the following considerations, we note that this expression is independent of $k_1, \dots, k_4$, $\ell_1, \dots, \ell_4$, and t and s. Also, we would like to use the notation that the expression is smaller in absolute value than κm for some $0 < \kappa < \infty$. Let further ϕ^* denote the adjoint operator of ϕ. Then, we obtain by using the definition of $\boldsymbol{U}_t$ and the aforementioned computation and following Aue et al. (2012) in applying the Cauchy-Schwarz inequality and Parseval's identity:

$$\begin{aligned}
&\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} A^{i_1}\Big(\mathbf{E}\Big[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^{\top}A^{i_2\top}A^{i_3}\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^{\top}\Big] \\
&\qquad\qquad -\mathbf{E}\left[\boldsymbol{U}_{t-i_1}\boldsymbol{U}_{t-i_2}^{\top}\right]A^{i_2\top}A^{i_3}\mathbf{E}\left[\boldsymbol{U}_{s-i_3}\boldsymbol{U}_{s-i_4}^{\top}\right]\Big)A^{i_4\top} \\
&= \sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d}\sum_{\ell_1=d+1}^{\infty}\sum_{\ell_2=d+1}^{\infty}\sum_{\ell_3=d+1}^{\infty}\sum_{\ell_4=d+1}^{\infty}\sum_{t=0}^{m}\sum_{s=0}^{m}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\sum_{i_3=0}^{\infty}\sum_{i_4=0}^{\infty} \\
&\quad A^{i_1}v_{k_1}v_{k_2}^{\top}A^{i_2\top}A^{i_3}v_{k_3}v_{k_4}^{\top}A^{i_4\top}\Big(\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right] \\
&\qquad -\mathbf{E}\left[\langle Y_{t-i_1-1}|v_{\ell_1}\rangle\langle Y_{t-i_2-1}|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_{s-i_3-1}|v_{\ell_3}\rangle\langle Y_{s-i_4-1}|v_{\ell_4}\rangle\right]\Big) \\
&\qquad\qquad \langle\phi(v_{\ell_1})|v_{k_1}\rangle\langle\phi(v_{\ell_2})|v_{k_2}\rangle\langle\phi(v_{\ell_3})|v_{k_3}\rangle\langle\phi(v_{\ell_4})|v_{k_4}\rangle \\
&\leqslant \kappa m\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d}\sum_{\ell_1=d+1}^{\infty}\sum_{\ell_2=d+1}^{\infty}\sum_{\ell_3=d+1}^{\infty}\sum_{\ell_4=d+1}^{\infty} \\
&\qquad |\langle\phi(v_{\ell_1})|v_{k_1}\rangle\langle\phi(v_{\ell_2})|v_{k_2}\rangle\langle\phi(v_{\ell_3})|v_{k_3}\rangle\langle\phi(v_{\ell_4})|v_{k_4}\rangle|\,||v_{k_1}||\,||v_{k_2}^{\top}|| \\
&\leqslant \kappa m\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d}\left(\sum_{\ell_1=d+1}^{\infty}\langle\phi(v_{\ell_1})|v_{k_1}\rangle^2\right)^{\frac{1}{2}}\left(\sum_{\ell_2=d+1}^{\infty}\langle\phi(v_{\ell_2})|v_{k_2}\rangle^2\right)^{\frac{1}{2}} \\
&\qquad \left(\sum_{\ell_3=d+1}^{\infty}\langle\phi(v_{\ell_3})|v_{k_3}\rangle^2\right)^{\frac{1}{2}}\left(\sum_{\ell_4=d+1}^{\infty}\langle\phi(v_{\ell_4})|v_{k_4}\rangle^2\right)^{\frac{1}{2}} \\
&\leqslant \kappa m\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d}\left(\sum_{\ell_1=d+1}^{\infty}\langle v_{\ell_1}|\phi^*(v_{k_1})\rangle^2\right)^{\frac{1}{2}}\left(\sum_{\ell_2=d+1}^{\infty}\langle v_{\ell_2}|\phi^*(v_{k_2})\rangle^2\right)^{\frac{1}{2}} \\
&\qquad \left(\sum_{\ell_3=d+1}^{\infty}\langle v_{\ell_3}|\phi^*(v_{k_3})\rangle^2\right)^{\frac{1}{2}}\left(\sum_{\ell_4=d+1}^{\infty}\langle v_{\ell_4}|\phi^*(v_{k_4})\rangle^2\right)^{\frac{1}{2}} \\
&= \kappa m\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\sum_{k_3=1}^{d}\sum_{k_4=1}^{d}||\phi^*(v_{k_1})||\,||\phi^*(v_{k_2})||\,||\phi^*(v_{k_3})||\,||\phi^*(v_{k_4})|| \\
&= \mathcal{O}(m).
\end{aligned}$$

□

Lemma 6.4.3. *Under the assumptions and notation of Section 6.2 it holds for $K_1, K_2 < \infty$ if $t_1 \leqslant t_2 \leqslant t_3 < t_4$*

$$\boldsymbol{E}\left[\langle Y_{t_1}|v_{\ell_1}\rangle\langle Y_{t_2}|v_{\ell_2}\rangle\langle Y_{t_3}|v_{\ell_3}\rangle\langle Y_{t_4}|v_{\ell_4}\rangle\right] = ||\phi||_{\mathcal{L}}^{t_4-t_3}K\nu_4(\delta_0)^4$$

and if $t \neq s$

$$\begin{aligned}
&\left(\boldsymbol{E}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]-\boldsymbol{E}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle\right]\boldsymbol{E}\left[\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\right)\\
&=||\phi||_{\mathcal{L}}^{|t-s|}K\nu_2(\delta_0)^2.
\end{aligned}$$

Proof. Without loss of generalization we assume $p=1$ for the proof and set for the first part $\tilde{Y}_{t_4}=\sum_{i=0}^{t_4-t_3+1}\phi^i\delta_{t-i}$. Then we obtain by using the fact that $\tilde{Y}_{t_4}$ is independent of Y_{t_1},Y_{t_2},Y_{t_3} and repeatedly applying the Cauchy-Schwarz inequality

$$\begin{aligned}
&\mathbf{E}\left[\langle Y_{t_1}|v_{\ell_1}\rangle\langle Y_{t_2}|v_{\ell_2}\rangle\langle Y_{t_3}|v_{\ell_3}\rangle\langle Y_{t_4}|v_{\ell_4}\rangle\right]\\
&\quad=\mathbf{E}\left[\langle Y_{t_4}-\tilde{Y}_{t_4}|v_{\ell_4}\rangle\langle Y_{t_1}|v_{\ell_1}\rangle\langle Y_{t_2}|v_{\ell_2}\rangle\langle Y_{t_3}|v_{\ell_3}\rangle\right]\\
&\quad\leqslant\mathbf{E}\left[\langle Y_{t_4}-\tilde{Y}_{t_4}|v_{\ell_4}\rangle^2\right]^{\frac{1}{2}}\mathbf{E}\left[\langle Y_{t_1}|v_{\ell_1}\rangle^2\langle Y_{t_2}|v_{\ell_2}\rangle^2\langle Y_{t_3}|v_{\ell_3}\rangle^2\right]^{\frac{1}{2}}\\
&\quad\leqslant\mathbf{E}\left[\int_0^1\left(Y_{t_4}-\tilde{Y}_{t_4}\right)^2(\tau)\mathrm{d}\tau\int_0^1 v_{\ell_4}^2(\tau)\mathrm{d}\tau\right]\mathbf{E}\left[\int_0^1 Y_{t_1}^2(\tau)\mathrm{d}\tau\int_0^1 Y_{t_2}^2(\tau)\mathrm{d}\tau\int_0^1 Y_{t_3}^2(\tau)\mathrm{d}\tau\right]\\
&\quad\leqslant\nu_2\left(Y_{t_4}-\tilde{Y}_{t_4}\right)(\nu_4(Y_0))^3\\
&\quad\leqslant\nu_2(\delta_0)||\phi||_{\mathcal{L}}^{t_4-t_3}K_1\left(\nu_4(Y_0)\right)^3\\
&\quad\leqslant\nu_4(\delta_0)^4||\phi||_{\mathcal{L}}^{t_4-t_3}K_1
\end{aligned}$$

For the second part we assume that $t>s$ and set $\tilde{Y}_t=\sum_{i=0}^{t-s+1}\phi^i\delta_{t-i}$ to obtain

$$\begin{aligned}
&\mathbf{E}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]-\mathbf{E}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\\
&\quad=\mathbf{Cov}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle,\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\\
&\quad=\mathbf{Cov}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle-\langle\tilde{Y}_t|v_{\ell_1}\rangle\langle\tilde{Y}_t|v_{\ell_2}\rangle,\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\\
&\quad=\mathbf{E}\left[\left(\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle-\langle\tilde{Y}_t|v_{\ell_1}\rangle\langle\tilde{Y}_t|v_{\ell_2}\rangle\right)\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\\
&\qquad\qquad-\mathbf{E}\left[\langle Y_t|v_{\ell_1}\rangle\langle Y_t|v_{\ell_2}\rangle-\langle\tilde{Y}_t|v_{\ell_1}\rangle\langle\tilde{Y}_t|v_{\ell_2}\rangle\right]\mathbf{E}\left[\langle Y_s|v_{\ell_3}\rangle\langle Y_s|v_{\ell_4}\rangle\right]\\
&\quad\leqslant||\phi||_{\mathcal{L}}^{|t-s|}K_2\nu_2(\delta_0)^2
\end{aligned}$$

by the same argumentation that was used before. The case $t<s$ follows in the same manner. □

Now, we state the proofs that were left out in Section 6.2.

Proof of Lemma 6.2.3. We note that

$$\begin{aligned}
\hat{A}^\top-A^\top&=B^{-1}G-A^\top\\
&=B^{-1}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_{t+1}-\bar{\boldsymbol{X}}\right)^\top-A^\top\\
&=B^{-1}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t-\bar{\boldsymbol{X}}\right)\left(A\left(\boldsymbol{X}_t-\bar{\boldsymbol{X}}\right)+\left(\boldsymbol{\mathcal{E}}_{t+1}-\bar{\boldsymbol{\mathcal{E}}}\right)\right)^\top-A^\top\\
&=B^{-1}\left(BA^\top+F\right)-A^\top\\
&=B^{-1}F.
\end{aligned}$$

From this equation we directly obtain the consistency of $\hat{A}$ by applying Lemma 6.4.1 and further

$$
\begin{aligned}
\hat{\Sigma}_q &= \frac{1}{m-q}\sum_{t=q+1}^{m}\left(\hat{\boldsymbol{\mathcal{E}}}_t - \bar{\boldsymbol{\mathcal{E}}}\right)\left(\hat{\boldsymbol{\mathcal{E}}}_t - \bar{\boldsymbol{\mathcal{E}}}\right)^\top \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\left(\boldsymbol{X}_t - \hat{A}\boldsymbol{X}_{t-1} - \left(\bar{\boldsymbol{X}} - \hat{A}\bar{\boldsymbol{X}}\right)\right)\left(\boldsymbol{X}_t - \hat{A}\boldsymbol{X}_{t-1} - \left(\bar{\boldsymbol{X}} - \hat{A}\bar{\boldsymbol{X}}\right)\right)^\top \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\left(\boldsymbol{\mathcal{E}}_t - \bar{\boldsymbol{\mathcal{E}}}_t - \left(\hat{A} - A\right)\left(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}}\right)\right)\left(\boldsymbol{\mathcal{E}}_t - \bar{\boldsymbol{\mathcal{E}}}_t - \left(\hat{A} - A\right)\left(\boldsymbol{X}_{t-1} - \bar{\boldsymbol{X}}\right)\right)^\top \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top + \left(\hat{A} - A\right)\frac{1}{m-q}\sum_{t=q}^{m-1}\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)^\top\left(\hat{A} - A\right)^\top \\
&\qquad + \frac{1}{m-q}\sum_{t=q}^{m-1}\left(\bar{\boldsymbol{\mathcal{E}}} - \boldsymbol{\mathcal{E}}_{t+1}\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)^\top\left(\hat{A} - A\right)^\top \\
&\qquad + \frac{1}{m-q}\sum_{t=q}^{m-1}\left(\hat{A} - A\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top + \frac{1}{m-q}\left(\hat{A} - A\right)B\left(\hat{A} - A\right)^\top - F^\top\left(\hat{A} - A\right)^\top + \left(\hat{A} - A\right)F \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top + \frac{1}{m-q}\left(\hat{A} - A\right)B\left(\hat{A} - A\right)^\top - F^\top B^{-1}F + F^\top {B^{-1}}^\top F \\
&= \frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top + \frac{1}{m-q}\left(\hat{A} - A\right)B\left(\hat{A} - A\right)^\top,
\end{aligned}
$$

We further have that

$$
\begin{aligned}
\frac{1}{m-q}\sum_{t=q+1}^{m}\boldsymbol{\mathcal{E}}_t\boldsymbol{\mathcal{E}}_t^\top &= \frac{1}{m-q}\sum_{t=q+1}^{m}\left(\boldsymbol{E}_t + \boldsymbol{U}_t\right)\left(\boldsymbol{E}_t + \boldsymbol{U}_t\right)^\top + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \Sigma_q + \mathcal{O}_P\left(\frac{1}{m} + \frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top &= \left(\bar{\boldsymbol{E}} + \bar{\boldsymbol{U}} + \frac{1}{\sqrt{n}}\bar{\boldsymbol{W}}\right)\left(\bar{\boldsymbol{E}} + \bar{\boldsymbol{U}} + \frac{1}{\sqrt{n}}\bar{\boldsymbol{W}}\right)^\top \\
&= \bar{\boldsymbol{E}}\bar{\boldsymbol{E}}^\top + \bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^\top + \bar{\boldsymbol{E}}\bar{\boldsymbol{U}}^\top + \bar{\boldsymbol{U}}\bar{\boldsymbol{E}}^\top + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \mathcal{O}_P\left(\frac{1}{m} + \frac{1}{m} + \frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

because easy calculations and Lemma 6.4.2 yield that

$$\begin{aligned}
\mathbf{E}\left[\bar{\boldsymbol{E}}\bar{\boldsymbol{E}}^\top\right] &= \mathcal{O}\left(\frac{1}{m}\right), \\
\mathbf{Var}\left[\bar{\boldsymbol{E}}\bar{\boldsymbol{E}}^\top\right] &= \mathcal{O}\left(\frac{1}{m^2}\right), \\
\mathbf{E}\left[\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^\top\right] &= \mathcal{O}\left(\frac{1}{m}\right), \\
\mathbf{Var}\left[\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^\top\right] &= \mathcal{O}\left(\frac{1}{m^2}\right).
\end{aligned}$$

Finally we obtain for the remaining term in Equation (6.11) by applying Lemma 6.4.1 and the fact that $\hat{A}^\top - A^\top = B^{-1}F$:

$$\begin{aligned}
&\left(\hat{A} - A\right) B \left(\hat{A} - A\right)^\top \\
&= F^\top B^{-1} B B^{-1} F \\
&= \left(\sum_{t=q}^{m-1} \left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top\right)^\top \left(\sum_{t=q}^{m-1} \left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)^\top\right)^{-1} \left(\sum_{t=q}^{m-1} \left(\boldsymbol{X}_t - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top\right) \\
&= \left(N_m^{-\frac{1}{2}} \sum_{t=q}^{m-1} \boldsymbol{Z}_t \left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top\right)^\top \left(N_m^{\frac{1}{2}\top} \left(\sum_{t=q}^{m-1} \boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right)^{-1} N_m^{\frac{1}{2}}\right) \left(N_m^{\frac{1}{2}} \sum_{t=q}^{n-1} \boldsymbol{Z}_t \left(\boldsymbol{\mathcal{E}}_{t+1} - \bar{\boldsymbol{\mathcal{E}}}\right)^\top\right) \\
&= V_m'^\top V_m V_m' \\
&= \mathcal{O}_P\left(1 + \frac{m}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right),
\end{aligned}$$

so that with

$$\hat{\Sigma}_q = \frac{1}{m-q} \sum_{t=q+1}^{m} \boldsymbol{\mathcal{E}}_t \boldsymbol{\mathcal{E}}_t^\top - \bar{\boldsymbol{\mathcal{E}}}\bar{\boldsymbol{\mathcal{E}}}^\top - \frac{1}{m-q}\left(\hat{A} - A\right) B \left(\hat{A} - A\right)^\top = \Sigma_q + \mathcal{O}_P\left(\frac{1}{m} + \frac{1}{\sqrt{n}}\right)$$

the second assertion follows to complete the proof. □

Proof of Lemma 6.2.4. We split the first d rows of A into $A_d = [A_1, A_2]$ and $\boldsymbol{X}_t^\top = [\boldsymbol{X}_{1,t}^\top, \boldsymbol{X}_{2,t}^\top]$, $\boldsymbol{X}_{1,t}$ and A_1 corresponding to the first q $\boldsymbol{Y}_t$, $\boldsymbol{X}_{2,t}$ and A_2 to the last $p-q$ $\boldsymbol{Y}_t$. Then, we can write the process as

$$\boldsymbol{Y}_t = A_1 \boldsymbol{X}_{1,t-1} + A_2 \boldsymbol{X}_{2,t-1} + \varepsilon_t. \tag{6.19}$$

Let $\hat{A}$ be the estimated matrix of fitting an AR(q) process to the data and $\tilde{A}$ the estimated matrix of an AR(p) fit. Both can be divided according to A. We assume further that $\hat{A}_1$ and $\hat{A}_2$ are ($\sqrt{m}$-consistent) estimators for A_1 and A_2 and set $H = B^{-1}$. Following the methods and argumentations of Paulsen (1984) (Proof of Theorem 1) and Anderson (2003)

(Chapter 8), we obtain

$$
\begin{aligned}
\hat{\Sigma}_q &= \frac{1}{m-k}\sum_{t=k+1}^{m}\left(\boldsymbol{Y}_t - \hat{A}_2\boldsymbol{X}_{2,t-1} - \hat{A}_1\boldsymbol{X}_{1,t-1}\right)\left(\boldsymbol{Y}_t - \hat{A}_2\boldsymbol{X}_{2,t-1} - \hat{A}_1\boldsymbol{X}_{1,t-1}\right)^\top \\
&= \frac{1}{m-k}\sum_{t=k+1}^{m}\left(\left(\boldsymbol{Y}_t - \tilde{A}\boldsymbol{X}\right) + \hat{A}_2\left(\boldsymbol{X}_{2,t-1} - B_{21}B_{11}^{-1}X_1\right)\right) \\
&\qquad\qquad\left(\left(\boldsymbol{Y}_t - \tilde{A}\boldsymbol{X}\right)^\top + \left(\boldsymbol{X}_{2,t-1} - B_{21}B_{11}^{-1}X_1\right)^\top\hat{A}_2^\top\right) \\
&= \frac{1}{m-k}\sum_{t=k+1}^{m}\Bigg(\left(\boldsymbol{Y}_t - \tilde{A}\boldsymbol{X}\right)\left(\boldsymbol{Y}_t - \tilde{A}\boldsymbol{X}\right)^\top \\
&\qquad\qquad + \hat{A}_2\left(\boldsymbol{X}_{2,t-1} - B_{21}B_{11}^{-1}X_1\right)\left(\boldsymbol{X}_{2,t-1} - B_{21}B_{11}^{-1}X_1\right)^\top\hat{A}_2^\top\Bigg) \\
&= \hat{\Sigma}_p + \frac{1}{m-k}\hat{A}_2\left(B_{22} - B_{21}B_{11}^{-1}B_{12}\right)\hat{A}_2^\top \\
&= \hat{\Sigma}_p + \hat{A}_2\underbrace{\frac{1}{m-q}H_{22}^{-1}}_{R}\hat{A}_2^\top \\
&= \hat{\Sigma}_p + A_2RA_2{}^\top + \mathcal{O}_P\left(\frac{1}{\sqrt{m}}\right)
\end{aligned}
$$

For $P = \begin{pmatrix} 0 \\ I \end{pmatrix}$ such that $H_{22} = P^\top B^{-1}P$ we have

$$
\begin{aligned}
R^{-1} = (m-q)H_{22} &= \sqrt{m-q}P^\top B^{-1}P\sqrt{m-q} \\
&= P^\top\sqrt{m-q}N_m^{-\frac{1}{2}\top}N_m^{\frac{1}{2}\top}\left(\sum_{t=1}^{m}\boldsymbol{X}_t\boldsymbol{X}_t^\top\right)^{-1}N_m^{\frac{1}{2}}N_m^{-\frac{1}{2}}\sqrt{m-q}P \\
&= P^\top\sqrt{m-q}N_m^{-\frac{1}{2}}VN_m^{-\frac{1}{2}}\sqrt{m-q}P
\end{aligned}
$$

We know that $N_m = \mathcal{O}_P(m)$, hence $\exists\,\eta' \in [0,1]$ such that $N_m^{-1} = \mathcal{O}_P(m^{-\eta'})$, hence, $mN_m^{-1} = \mathcal{O}_P(m^{-\eta'+1})$. In addtion, $V = \mathcal{O}_P(1)$. Thus, $\exists\,\eta \in [0,1]$ such that $R^{-1} = \mathcal{O}_P(m^{-\eta+1})$.
Further, R is positive definite and thus $\alpha R\alpha^\top \geqslant \lambda||\alpha||^2\ \forall\,\alpha \in \mathbb{R}^{(p-q)d\times(p-q)d}$, where λ is the smallest eigenvalue of R. This implies that $\frac{1}{\lambda}$ is the largest eigenvalue of R^{-1}, that is also positive definite. Since $\frac{1}{\lambda} = \sup_{||\alpha||=1}||R^{-1}\alpha|| = \mathcal{O}_P(m^{-\eta+1})$ it follows that $\lambda = \mathcal{O}_P(m^{\eta-1})$ and $\alpha R\alpha = \mathcal{O}_P(m^{\eta-1})$ and $0 < A_2RA_2{}^\top = \mathcal{O}_P(m^{\eta-1})$ if $A_2 \neq 0$, what is fulfilled by assumption. □

Acknowledgements

First of all, I want to express my gratitude to my advisor Jens-Peter Kreiß for offering me a position in in his group, ensuring a pleasant work environment and giving me the opportunity to write this thesis. In particular, I would like to thank him for the confidence he placed in me and my work, whether research or teaching, and the possibilities he gave me to pursue the things I was interested in.

Further, I would like to thank Alexander Aue for being a very kind host during a research visit in Davis making this visit delightful and introducing me into a new subject providing possibilities for successful research. I also want to thank Alexander Aue and Thomas Lee for being trusting coauthors.

Efstathios Paparodits agreed to act as a referee for this thesis, what I highly appreciate.

Further thanks go to Alexander Lindner for the kind collaboration inside our department and my other collegues for the pleasant atmosphere in our department and also to our students who showed interest in our subject and provided a nice teaching atmosphere.

Last but not least I want to thank my family and my friends for their constant support and many times of distraction and who made this time a really great time.

Bibliography

Akaike, H. (1969): *Fitting Autoregressive Models for Prediction,* The Annals of the Institute of Statistical Mathematics 21, 243-247.

Akaike, H. (1974): *A new Look at the Statistical Model Identification,* IEEE Transactions on Automatic Control 19, 716-723.

Akaike, H. (1978): *Time Series Analysis and Control through Parametric Models.* In D.F. Findely, editor, Applied Time Series Analysis, New York Academic Press.

Andel, J. (1976): *Autoregressive series with random parameters,* Math. Operationsforsch. Statist. 7: 5, 735 - 741.

Anderson, T.W. (2003): *An Introduction to Multivariate Statistical Analysis,* Wiley.

Aue, A. (2003): *Sequential Change–Point Analysis based on Invariance Principles,* Dissertation Universität Köln.

Aue, A. (2006): *Testing for parameter stability in RCA(1) time series,* Journal of Statistical Planning and Inference 136, 3070-3089.

Aue, A. and Horvath, L. (2011): *Quasi-Likelihood Estimation in Stationary and Nonstationary Autoregressive Models with Random Coefficients,* Statistica Sinica 21, 973-999.

Aue, A., Horvath, L., and Steinebach, J. (2006): *Estimation in Random Coefficient Autoregressive Models,* Journal of Time Series Analysis 27, 61-76.

Aue, A., Norinho, D., and Hörmann, S. (2012): *On the predicion of functional time series,* preprint.

Aue, A., Fink, T., and Lee, T. (2013): *Order Determination of Functional Autoregressive Processes,* preprint UCD Davis, TU Braunschweig.

Berkes, I., Horvath, L., and Kokoszka, P. (2003): *GARCH processes: structure and estimation,* Bernoulli 9, 201-227.

Berkes, I., Horvath, L., and Ling, S. (2009): *Estimation in Nonstationary Random Coefficient Autoregressive Models,* Journal of Time Series Analysis 30, 395-416.

Berkes, I., Gabrys, R., Horvath, L., and Kokoszka, P. (2009): *Detecting Changes in the Mean of Functional Observations,* Journal of the Royal Statistical Society 71, 927-946.

Billingsley, P. (1968): *Convergence of Probability Measures,* John Wiley & Sons.

Bosq, D. (2000): *Linear Processes in Function Spaces,* Springer.

Bosq, D. and Blanke, D. (2007): *Inference and Prediction in Large Dimensions,* Wiley.

Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods,* Springer, New York.

Bühlmann, P. (1997): *Sieve Bootstrap for Time Series,* Bernoulli 3, No. 2, 123-148.

Carlstein, E. (1986): *The use of subseries methods for estimating the variance of a general statistic from a stationary time series,* Annals of Statistics 14, 1171-1179.

Dahlhaus, R. and Janas, D. (1996): *A frequency domain bootstrap for ratio staistics in time series analysis,* Annals of Statistics 24, 1934-1963.

Dauxois, J., Pousse, A., and Romain, Y. (1982): *Asymptotic theory for principal component analysis of a vector random function,* Journal of Multivariate Analysis 12, 136-154.

Dedecker, J., Doukhan, P., Lang, G., Rafael, J., Louhichi, S., Prieur, C. (2007): *Weak Dependence,* Springer Lecture Notes in Statistics 190.

Efron, B. (1979): *Bootstrap Methods: Another Look at the Jackknife,* Annals of Statistics 7, 1-26.

Efron, B. and Tibshirani, R. (1986): *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy,* Statistical Science 1, 54-77.

Elstrodt, J. (2009): *Maß und Integrationstheorie, 6. Auflage,* Springer.

Fan, J. (1991): *Asymptotic Normality for Deconvolution Kernel Density Estimators,* Indian Journal of Statistics 53, 97-110.

Feigin, P.D. and Tweedie, R.L.(1985): *Random Coefficient Autoregressive Processes. A Markov Chain Analysis of Stationarity anf Finiteness of Moments,* Journal of Time Series Analysis 6, 1-14.

Ferraty, F. and Romain, Y. (eds) (2011): *The Oxford Handbook of Functional Data Analysis,* Oxford Press.

Ferraty, F. and Vieu, P. (2006): *Nonparametric Functional Data Analysis: theory and practise,* Springer.

Fink, T. and Kreiß, J.-P. (2012): *Bootstrap for Random Coefficient Autoregressive Models,* submitted.

Fink, T. and Kreiß, J.-P. (2013): *Simultaneous Bootstrap for all three Parameters in Random Coefficient Autoregressive Models,* preprint TU Braunschweig.

Franke, J. and Härdle, W. (1992): *On Bootstrapping Kernal Spectral Estimates,* Annals of Statistics 20, 121-145.

Freedman, D. (1984): *On Bootstrapping two-stage least-squares Estimates in Stationary Linear Models,* The Annals of Statistics 12, 827-842.

Hall, P., Müller, H., and Wang, J. (2006): *Properties of Principal Component Methods for Functional and Longitudinal Data Analysis,* The Annals of Statistics 34, 1493-1517.

Hannan, E.J. and Quinn, B.G. (1979): *The Determination of the Order of an Autoregression,* Journal of the Royal Statistical Society Series B 41, 190-195.

Härdle, W., Horowitz, J., and Kreiß, J.-P. (2003): *Bootstrap Methods for Time Series,* International Statistical Review 71, 435-459.

Hörmann, S. and Kokoszka, P. (2010): *Weakly dependent functional data,* The Annals of Statistics 38, 1845-1884.

Hurvich, C.M. and Tsai, C.L. (1989): *Regression and Time Series Model Selection in small Samples,* Biometrika 76, 297-307.

Jentsch, C. and Kreiss, J.-P. (2010): *The Multiple Hybrid Bootstrap - Resampling Multivariate Linear Processes,* Journal of Multivariate Analysis 101, 2320-2345.

Kirch, C. and Politis, D. (2011): *TFT-Bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain,* Annals of Statistics 39, 1427-1470.

Klenke, A. (2006): *Wahrscheinlichkeitstheorie,* Springer.

Kokoszka, P. and Reimherr, M. (2012): *Determining the Order of the Functional Autoregressive Model,* Journal of Time Series Analysis.

Koul, H. and Schick, A. (1996): *Adaptive Estimation in a Random Coefficient Autoregressive Model,* Annals of Statistics 24, 1025-1052.

Kreiss, J.-P. (1988): *Asymptotical Inference for a Class of Stochastic Processes,* Habilitationsschrift, Universität Hamburg.

Kreiss, J.-P. (1997): *Asymptotical Properties of Residual Bootstrap for Autoregressions,* Manuscript, TU Braunschweig.

Kreiss, J.P. and Lahiri, S.N. (2012): *Bootstrap Methods for Time Series.* In: Handbook of Statistics, Vol 30: Time Series-Methods and Applications, Elsevier-North Holland, 3-26.

Kreiss, J.-P. and Neuhaus, G. (2006): *Einführung in die Zeitreihenanalyse,* Springer.

Kreiss, J.-P. and Paparoditis, E. (2003): *Autoregressive aided periodogram bootstrap,* Annals of Statistics 31, 1923-1955.

Kreiss, J.-P. and Paparoditis, E. (2011): *Bootstrap for Dependent Data: A Review (with Discussion),* Journal of the Korean Statistical Society 40, 357-395.

Kreiss, J.-P. and Paparoditis, E. (2012): *The Hybrid Wild Bootstrap for Time Series,* Journal of the American Statistical Association 107, 1073-1084.

Kreiss, J.-P., Paparoditis, E., and Politis, D.N. (2011): *The Range of Vailidity of the Autoregressive Sieve Bootstrap,* The Annals of Statistics 39, 2103-2130.

Künsch, H.R. (1989): *The Jacknife and the Bootstrap for General Stationary Observations,* Annals of Statistics 17, 1217-1241.

Lahiri, S.N. (2003): *Resampling Methods for Dependent Data,* Springer, New York.

Lee, T.C.M. (2001): *An Introduction to Coding Theory and the Two-Part Minimum Description Length Principle,* International Statistical Review 69, 169-183.

Lohse, K. (1987): *Consistency of the Bootstrap,* Statistics and Decisions 5, 353-363.

Lütkepohl, H. (2005): *A New Introduction to Multiple Time Series Analysis,* Springer.

Neumann, M.H. and Paparoditis, E. (2008): *Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap approximations,* Bernoulli 14, 14-46.

Nicholls, D.F. and Quinn, B.G. (1980): *The Estimation of Random Coefficient Autoregressive Models I,* Journal of Time Series Analysis 1, 37-46.

Nicholls, D.F. and Quinn, B.G. (1982): *Random Coefficient Autoregressive Models: An Introduction,* Springer Lecture Notes in Statistics.

Paulsen J. (1984): *Order Determination of Multivariate Autoregressive Time Series with Unit Roots,* Journal of Time Series Analysis.

Pham, D.T. (1986): *The Mixing Property of Bilinear and Generatlized Random Coefficient Autoregressive Models,* Stochastic Processes and their Applications 23, 291-300.

Politis, D.N. and Romano, J.P. (1994): *The Stationary Bootstrap,* Journal of the American Statistical Association 89, 1303-1313.

Praskova, Z. (2003): *Wild bootstrap in RCA(1) model,* Kybernetica 29, 1-12.

Quinn, B.G. (1980): *Order Determination for a Multivariate Autoregression,* Journal of the Royal Statistical Society Series B 42, 182-185.

Ramsey, J.O. and Silverman, B.W. (2005): *Functional Data Analysis,* Springer.

Reinsel, G.C. (2003): *Elements of Multivariate Time Series Analysis, second edition,* Springer.

Rissanen, J. (1989): *Stochastic Complexity in Statistical Inquiry,* World Scientific, Singapore.

Rissanen, J. (2007): *Information and Complexity in Statistical Modeling,* Springer.

Robinson, P. (1978): *Statistical Inference for a Random Coefficient Autoregressive Model,* Scandinavian Journal of Statistics 5, 163-168.

Rosenblatt, M. (1956): *Remarks on Some Nonparametric Estimates of a Density Function,* Annals of Mathematical Statistics 27, 832-837.

Rosenblatt, M. (1971): *Markov Processes. Structure and Asymptotic Behavior,* Springer Berlin.

Schick, A. (1996): *$\sqrt{n}$-consistent Estimation in a Random Coefficient Autoregressive Model,* Australian & New Zealand Journal of Statistics 38, 155-160.

Shao, X. and Wu, W.B. (2007): *Asymptotic spectral theory for nonlinear time series,* Annals of Statistics 35, 1173-1801.

Stadtlober, E., Hörmann, S., and Pfeiler, B. (2008): *Quality and performance of a PM10 daily forecasting model,* Athmospheric Environment 42, 1098-1109.

Tiao, G.C. and Tsay, R.S. (1989): *Model Specification in Multivariate Time Series (with discussion),* Journal of the Royal Statistical Society Series B 51, 157-213.

Tong, H. (1990): *Non-linear Time Series: A Dynamical System Approach,* Oxford: Clarendon Press.